

GenCore version 5.1.3

Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 08:52:30 : Search time 958 Seconds
(without alignments)
327.660 Million cell updates/sec

Title: US-09-820-203A-1

Perfect score: 15

Sequence: 1 agtaggggaatgag 15

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_hgt:*

3: gb_in:*

4: gb_ov:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_ov:*

21: em_ov:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_to:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_hgt_hum:*

31: em_hgt_inv:*

32: em_hgt_other:*

33: em_hgtg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match Length	DB ID	Description

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	JOURNAL	COMMENT
1	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	Contact: Richard M. Myers Stanford Human Genome Center (SHGC) Stanford University School of Medicine Department of Genetics, M-344, Stanford, CA 94305, USA Tel: 4157259687 Fax: 4157259689 Email: myers@shgc.stanford.edu
2	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
3	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
4	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
5	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
6	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
7	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
8	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
9	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
10	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
11	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
12	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
13	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
14	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
15	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
16	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
17	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
18	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
19	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
20	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
21	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
22	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
23	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
24	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
25	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
26	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
27	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
28	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
29	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
30	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
31	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
32	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
33	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
34	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
35	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
36	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
37	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
38	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
39	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
40	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
41	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
42	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
43	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
44	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	
45	GI5398	human STS SHGC-16139, sequence tagged site.	GI5398	GI:1132161	STS; STS sequence; primer; sequence tagged site.	human.	Homo sapiens	1 (bases 1 to 250)	Myers,R.M.	Unpublished	

ALIGNMENTS

GI5398 human STS SHGC-16139, sequence tagged site. 250 bp DNA linear STS 04-JAN-1996

GI5398 human STS SHGC-16139, sequence tagged site.

GI5398.1 GI:1132161

STS; STS sequence; primer; sequence tagged site.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Myers,R.M.

Unpublished

1 (bases 1 to 250)

Myers,R.M.

Unpublished

1 (bases 1 to 250)

Myers,R.M.

Unpublished

1 (bases 1 to 250)

Myers,R.M.

Unpublished

1 (bases 1 to 250)

Myers,R.M.

Unpublished

1 (bases 1 to 250)

Primer A: AGCAATTGGTAAGTGGCTGG
Primer B: TGTACGCTGACTTTCCT
STS size: 141
PCR Profile:

Initial incubation: 94 degrees C for 90 seconds

Denaturation: 94 degrees C for 15 seconds
Annealing: 62 degrees C for 23 seconds
Polymerization: 72 degrees C for 30 seconds
PCR Cycles: 30

Thermal Cycler: Perkin Elmer 9600

Protocol:

Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Taq Polymerase: 0.05 units/ul
Total Vol: 10 ul

Buffer:

MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 20 mM
pH: 8.3

Prepared with primer pairs derived from T77464--Merck/UniEST.

FEATURES

Location/Qualifiers
1..250
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="8"
17..157
primer_bind
17..36
primer_bind
complement(138..157)
69 t
BASE COUNT
91 a 27 c 63 g 69 t
ORIGIN

Query Match 100.0%; Score 15; DB 11; Length 250;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
|||||
Db 115 AGTATGGGAATGAG 129

RESULT 2

AY072077/c

LOCUS

AY072077 1422 bp mRNA linear PLN 16-JAN-2002
Arabidopsis thaliana putative Myb-related transcription activator
protein (At5g47390) mRNA, complete cds.

ACCESSION

AY072077

VERSION

FLI_CDNA

KEYWORDS

thale cress.

SOURCE

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 1422)

Yamada,K., Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M.,

Deng,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L.,

Tang,C.C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L.,

Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,

Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,

Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,

Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A.,

Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.

Full length cDNA of gene At5g47390 (GI:15238083)

Unpublished

2 (bases 1 to 1422)

Yamada,K., Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M.,

Deng,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L.,

Tang,C.C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L.,

TITLE

JOURNAL

COMMENT

Submitted (02-JAN-2002) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN
Arabidopsis Full-length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the
sequencing and annotation of the RAPL cDNAs: Yamada,K., Banh,J.,
Chan,M.M., Chang,C.H., Chang,E., Dale,J.M., Deng,J.M.,
Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C.,
Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L., Chen,H.,
Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J.,
Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J., Shinn,P.,
Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
contributed equally to this work as PIs.

FEATURES

source

Location/Qualifiers

1..1422

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/chromosome="5"

/clone="RAPL09-94-B16 (R13763)"

/note="This clone is in a modified pBluescript vector
(FIC-1) as a BamHI/XhoI insert.

ecotype: Columbia"

1..1422

/gene="At5g47390"

1..157

/gene="At5g47390"

158..1255

/gene="At5g47390"

/codon_start=1

/evidence=experimental

/product="putative Myb-related transcription activator
protein"

/protein_id="AAL59900.1"

/db_xref="GI:18175632"

/translation="MTRRCSCHNHGNSRTPCNRGVKLFQVRLTKSIRKSASMINL
SHYTGSGSGHGTGNTSPGPDVDPHVGAGYSEDFVAGSSSRERKKKGTPTWTFEE
HMFLLGLKLGKSDMEGISRYVTRPTQVASHAQKYFIROSNSRKRKRSI.FDM
VFEDVGDIPMDLQPEEDNIPVETEMGDSHQTLAPSLHAPSLIEECESMST
NSTGTETATAAASSSRLEETLOLOPQPOLPGSPFTIYPTVFPYPPFPPI
WPAGYVPEPKKEETHELTPAVHSKAPINVDLLGMSKLSLAESNKHGFSQSLSL
KLGGSSSRQSAFHPNPSSDSDIKSVTHAL"

1256..1422

/gene="At5g47390"

misc_difference 1407

/gene="At5g47390"

/note="not present in genomic sequence"

BASE COUNT 389 a 332 c 311 g 390 t

ORIGIN

Query Match 100.0%; Score 15; DB 8; Length 1422;

Best Local Similarity 100.0%; Pred. No. 7e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15

|||||

Db 939 AGTATGGGAATGAG 925

|||||

RESULT 3

HSTIF2GEN/c

LOCUS HSTIF2GEN 6156 bp mRNA linear PRI 10-MAR-1997
 DEFINITION H.sapiens mRNA for transcriptional intermediary factor 2.
 ACCESSION X97674.1 GI:1877214
 VERSION X97674.1
 KEYWORDS alternatively spliced; nuclear receptor coactivator; TIF2 gene; transcriptional mediator.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 6156)
 AUTHORS Voegel J.J., Heine M.J., Zechel C., Chambon P. and Gronemeyer H.
 TITLE TIF2, a 160 kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors
 JOURNAL EMBO J. 15 (14), 3667-3675 (1996)
 MEDLINE 96312964
 REFERENCE 2 (bases 1 to 6156)
 AUTHORS Voegel J.J.
 TITLE Direct Submission
 JOURNAL Submitted (22-APR-1996) J.J. Voegel, IGMC Inst.de Genet.et Biol.Mol.et Cell., CNRS-INSERM-Univ.Louis Pasteur, B.P.163, C.U. de Strasbourg, F-67404 ILLKIRCH CEDEX, FRANCE
 REVISION Revisé by author 25-JUL-96 and 10-MAR-97
 COMMENT On Mar 11, 1997 this sequence version replaced gi:1490315.
 FEATURES
 Location/Qualifiers
 1..6156
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="placenta"
 /clone_lib="lambdaEX10x Ref.No.56"
 163..4557
 /gene="TIF2"
 163..4557
 /gene="TIF2"
 /function="transcriptional mediator for ligand-dependent activation function AF-2 of nuclear receptors"
 /codon_start=1
 /product="transcriptional intermediary factor 2"
 /protein_id="CAA66263.1"
 /db_xref="GI:1877215"
 /db_xref="SPTREMBL:Q15596"
 /translation="MSGMGENTSPRAETRRKKECPDQGLSPKRNTEKRNQENK YIEELAEIPANFDINDENFKPKCAILKETVKQIRQIKQERAAANIDEVOKSDV SSTQCQVIXDKALGPMLEALDGEFFVNLGNVVFVSENVOYLRVNOELMKNKSV SILVGDHTEFYKNLLPKSTVINGSGEPKRNHSTFNCMLVKPLPDPSEEGHDNQ EAKQYETMOCFAVSOPKSTKEGEDLQCLICVAPRPMKRPVLPSPSEFTTRQDL QKITSLDSTMRAMKFEWEDLVKRCIQKFAHQHGESVYAKRHHEVLRLGLAFS QYFRSLSDGTLVAAQTKSLIRSQTNPEPLVLSLHMLHREONVYVNPDLTGTMG KPLNPLSSNPAHQALCSGNGPQDMTLLSNIFFNPNGKEQMGPMGRFGSGGMHV SGMQATTPQGSNVALKNSPQSGNNGNPGQPTSLSPRMRSPGVAGSPRIIPPSQFS PAGLSHPGVGVCSSGNSHSTNSSNALQALSGHVSLSGLSLASPLDKMGLNSP VNNPPLSKMGLSDCKTGLYGEPESEGTGQAESCHPGCEKETNDPNLNPVAVSE RADQSRHDSKGTQKLLQTLTKSDQMEFSPPLASLSDTNKSTGLSPGSGTHGTS LKSKHLRLHQDSSPDLAKLTATATCKDLSQESSSTAPGEVTKQEPYSPKK ENALLYLKDDTKDIDGLPEITPKLRDLSDTPASNTKLIAMKTEKMSPEPDQ PGSELNLEILDLQNSOLPDPDTRCAPAGSVYKQALINDLMOLTAENSPVTPV GAKTALRSOSTFNNRPGQLGRLPLNQNLFDITLQSPGAGFPFPIRNSPSYPIV PQQPMNGMGIGNUGNLSGMSGMSASRPTMPSGEWAPQSSAVRVCAATTSAM NRPVQGMGIRNPAISIPMRPSSQPGQRTLQSQVKNIGPSELMNMGPGYQQQAAP NOTAPWESILP IDQAFASQNRQPGSSFDLLCPHPAESPDEGALLDLYLALR NFDGLBEDILALGIPELVVSQVADPEQFSSQSDSNIMLEOKAPVFPQCYASQAQAG SVSPMDQPNFTMGORPSYATLRMPQGLRPTGLVQNPQLRLQLRHLQHQNRQ PLMNQISNVNVNLTLRGVPYTPANMAOLQRORELLNHLRQRMQOQOQOORT LMRGGLNTPSWAPSGMPATMSNPRIPQANAQFFPPPNYISQQPDGPTCAIT POSPLSPRAHTSPXAKQSQANPAYQAPSDINGWAGQNGMSQSPHPHCOQ ANTSMYNNMINNTATNTGMSNMNQMTQISMTSVTSVPTSLGSSMGPEQVNDPA LRGNLFPNOLPGMDMKQEGDTRKYC"
 misc_feature 2768..2974
 /gene="TIF2"
 /note="putative alternatively spliced region"
 BASE COUNT 1799 a 1493 c 1406 g 1458 t
 ORIGIN

Query Match 100.0% Score 15; DB 9; Length 6156;
 Best Local Similarity 100.0%; Pred. No. 4.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGTATGGGAATGAG 15
 Db 5420 AGTATGGGAATGAG 5406
 RESULT 4
 AC090546/c
 LOCUS
 DEFINITION
 AC090546
 SEQUENCE SAMPLING.
 AC090546.1 GI:13184100
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 62305)
 AUTHORS Birren B., Linton L., Nusbaum C. and Lander E.
 TITLE Homo sapiens chromosome 15, clone RP11-552E10
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 62305)
 AUTHORS Birren B., Linton L., Nusbaum C., Lander E., Allen N., Anderson S., Barrera N., Bastien V., Boguslavsky L., Bonkhallo B., Brown A., Camarata J., Campopiano A., Choepel Y., Collangelo M., Collins S., Collymore A., Cooke P., DeArrellano K., Dewar K., Diaz J.S., Dodge S., Faro S., Ferreira P., FitzHugh W., Gage D., Galagan J., Gardyna S., Ginde S., Goyette M., Graham L., Grand-Pierre N., Hago S., Heaford A., Horton L., Hulme W., Iliev I., Johnson K., Jones C., Karats A., Lakocque K., Lamazares R., Landers T., Leoczky J., Levine R., Liu G., Maclean C., Macdonald P., Marquis N., Matthews C., McCarthy M., McEwan P., McKernan K., McPheeters R., Meldrum J., Meneus L., Mihova T., Mlenaga V., Murphy T., Naylor J., Nguyen C., Norbu C., Norman C.H., O'Connor T., O'Donnell P., O'Neill D., Oliver J., Peterson K., Phukhang P., Pierre N., Pollara V., Raymond C., Reilly R., Rieback M., Riley R., Rise C., Rogov P., Roman J., Rosetti M., Roy A., Santos R., Schauer S., Schupbach R., Seaman S., Severy P., Soune C., Spencer B., Stange-Thomann N., Stojanovic N., Strauss N., Subramanian A., Talamas J., Testave S., Theodorou J., Travers M., Travis N., Triggillo J., Vassiliev H., Viel R., Vo A., Wilson B., Wu X., Wyman D., Ye W.J., Young G., Zaimoun J., Zembek L., Zimmer A. and Zody M.
 Direct Submission
 Submitted (02-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Institute
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L22431
 Center clone name: 552_E10

 * NOTE: This record contains 77 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will

* be preserved. 709: contig of 709 bp in length
1 710 809: gap of 100 bp
810 1508: contig of 699 bp in length
1509 1608: gap of 100 bp
1609 2319: contig of 711 bp in length
2320 2419: gap of 100 bp
2420 3141: contig of 722 bp in length
3142 3241: gap of 100 bp
3242 3935: contig of 694 bp in length
3936 4035: gap of 100 bp
4036 4736: contig of 701 bp in length
4737 4836: gap of 100 bp
4837 5548: contig of 712 bp in length
5549 5648: gap of 100 bp
5649 6361: contig of 713 bp in length
6362 6461: gap of 100 bp
6462 7176: contig of 715 bp in length
7177 7276: gap of 100 bp
7277 7985: contig of 709 bp in length
7986 8085: gap of 100 bp
8086 8788: contig of 703 bp in length
8789 8888: gap of 100 bp
8889 9570: contig of 682 bp in length
9571 9670: gap of 100 bp
9671 10391: contig of 721 bp in length
10392 10491: gap of 100 bp
10492 11194: contig of 703 bp in length
11195 11294: gap of 100 bp
11295 12009: contig of 715 bp in length
12010 12109: gap of 100 bp
12110 12821: contig of 712 bp in length
12822 12921: gap of 100 bp
12922 13622: contig of 701 bp in length
13623 13722: gap of 100 bp
13723 14425: contig of 703 bp in length
14426 14525: gap of 100 bp
14526 15254: contig of 729 bp in length
15255 15354: gap of 100 bp
15355 16075: contig of 721 bp in length
16076 16175: gap of 100 bp
16176 16880: contig of 705 bp in length
16881 16980: gap of 100 bp
16981 17695: contig of 715 bp in length
17696 17795: gap of 100 bp
17796 18495: contig of 700 bp in length
18496 18595: gap of 100 bp
18596 19297: contig of 702 bp in length
19298 19397: gap of 100 bp
19398 20113: contig of 716 bp in length
20114 20213: gap of 100 bp
20214 20946: contig of 733 bp in length
20947 21046: gap of 100 bp
21047 21769: contig of 723 bp in length
21770 21869: gap of 100 bp
21870 22577: contig of 708 bp in length
22578 22677: gap of 100 bp
22678 23383: contig of 706 bp in length
23384 23483: gap of 100 bp
23484 24178: contig of 695 bp in length
24179 24278: gap of 100 bp
24279 24986: contig of 708 bp in length
24987 25086: gap of 100 bp
25087 25797: contig of 711 bp in length
25798 25897: gap of 100 bp
25898 26603: contig of 706 bp in length
26604 26703: gap of 100 bp
26704 27406: contig of 703 bp in length
27407 27506: gap of 100 bp
27507 28220: contig of 714 bp in length
28221 28320: gap of 100 bp
28321 29026: contig of 706 bp in length
29027 29126: gap of 100 bp

* 29127 29840: contig of 714 bp in length
29841 29940: gap of 100 bp
29941 30666: contig of 726 bp in length
30667 30766: gap of 100 bp
30767 31469: contig of 703 bp in length
31470 31569: gap of 100 bp
31570 32280: contig of 711 bp in length
32281 32380: gap of 100 bp
32381 33094: contig of 714 bp in length
33095 33194: gap of 100 bp
33195 33886: contig of 692 bp in length
33887 33986: gap of 100 bp
33987 34702: contig of 716 bp in length
34703 34802: gap of 100 bp
34803 35511: contig of 709 bp in length
35512 35611: gap of 100 bp
35612 36322: contig of 711 bp in length
36323 36422: gap of 100 bp
36423 37124: contig of 702 bp in length
37125 37224: gap of 100 bp
37225 37940: contig of 716 bp in length
37941 38040: gap of 100 bp
38041 38735: contig of 695 bp in length
38736 38835: gap of 100 bp
38836 39547: contig of 712 bp in length
39548 39647: gap of 100 bp
39648 40357: contig of 710 bp in length
40358 40457: gap of 100 bp
40458 41158: contig of 701 bp in length
41159 41258: gap of 100 bp
41259 41966: contig of 708 bp in length
41967 42066: gap of 100 bp
42067 42769: contig of 703 bp in length
42770 42869: gap of 100 bp
42870 43591: contig of 722 bp in length
43592 43691: gap of 100 bp
43692 44426: contig of 735 bp in length
44427 44526: gap of 100 bp
44527 45247: contig of 721 bp in length
45248 45347: gap of 100 bp
45348 46056: contig of 709 bp in length
46057 46156: gap of 100 bp
46157 46857: contig of 701 bp in length
46858 46957: gap of 100 bp
46958 47640: contig of 683 bp in length
47641 47740: gap of 100 bp
47741 48453: contig of 713 bp in length
48454 48553: gap of 100 bp
48554 49272: contig of 719 bp in length
49273 49372: gap of 100 bp
49373 50093: contig of 721 bp in length
50094 50193: gap of 100 bp
50194 50884: contig of 691 bp in length
50885 50984: gap of 100 bp
50985 51693: contig of 709 bp in length
51694 51793: gap of 100 bp
51794 52505: contig of 712 bp in length
52506 52605: gap of 100 bp
52606 53318: contig of 713 bp in length
53319 53418: gap of 100 bp
53419 54134: contig of 716 bp in length
54135 54234: gap of 100 bp
54235 54966: contig of 732 bp in length
54967 55066: gap of 100 bp
55067 55774: contig of 708 bp in length
55775 55874: gap of 100 bp

Query Match 100.0% Score 15; DB 2; Length 62405;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGATGAG 15

|||||

DB 46546 AGTATGGGAATGAG 46532

RESULT 5

AC100883

LOCUS

DEFINITION

AC100883

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC100883 64671 bp DNA linear HTG 23-NOV-2001
Mus musculus clone RP23-67J9, LOW-PASS SEQUENCE SAMPLING.

AC100883.1 GI:17059657

HTG; HTGS-PHASE0.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 64671)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

2 (bases 1 to 64671)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,

Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,

Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,

Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,

Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,

Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,

Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,

Jones,C., Kanat,A., Karatas,A., Kellis,C., LaRocque,K.,

Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,

MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,

McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meidrim,J.,

Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,

Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,

Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,

Ramond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,

Seaman,S., Severy,P., Spencer,B., Santos,R., Schauer,S., Schupback,R.,

Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L14493

Center clone name: 67_J_9

* NOTE: This record contains 77 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

* 1 740: contig of 740 bp in length

* 741 840: gap of 100 bp

* 841 1599: contig of 759 bp in length

* 1600 1699: gap of 100 bp

* 1700 2429: contig of 730 bp in length

* 2430 2529: gap of 100 bp

* 3250 3351: contig of 722 bp in length

* 3252 3351: gap of 100 bp

* 3352 4070: contig of 719 bp in length

* 4071 4170: gap of 100 bp

* 4171 4910: contig of 740 bp in length

* 4911 5010: gap of 100 bp

* 5011 5760: contig of 750 bp in length

* 5761 58 0: gap of 100 bp

* 5861 6607: contig of 747 bp in length

* 6608 6707: gap of 100 bp

* 6708 7432: contig of 725 bp in length

* 7433 7532: gap of 100 bp

* 7533 8289: contig of 757 bp in length

* 8290 8389: gap of 100 bp

* 8390 9132: contig of 743 bp in length

* 9133 9232: gap of 100 bp

* 9233 9964: contig of 732 bp in length

* 9965 10064: gap of 100 bp

* 10065 10762: contig of 698 bp in length

* 10763 10962: gap of 100 bp

* 10963 11614: contig of 752 bp in length

* 11615 11714: gap of 100 bp

* 11715 12456: contig of 742 bp in length

* 12457 12556: gap of 100 bp

* 12557 13301: contig of 745 bp in length

* 13302 13401: gap of 100 bp

* 13402 14118: contig of 717 bp in length

* 14119 14218: gap of 100 bp

* 14219 14978: contig of 760 bp in length

* 14979 15078: gap of 100 bp

* 15079 15855: contig of 777 bp in length

* 15856 15955: gap of 100 bp

* 15956 16673: contig of 720 bp in length

* 16676 16775: gap of 100 bp

* 16776 17505: contig of 730 bp in length

* 17506 17605: gap of 100 bp

* 17606 18338: contig of 733 bp in length

* 18339 18438: gap of 100 bp

* 18439 19159: contig of 721 bp in length

* 19160 19259: gap of 100 bp

* 19260 20007: contig of 748 bp in length

* 20008 20107: gap of 100 bp

* 20108 20860: contig of 753 bp in length

* 20861 20960: gap of 100 bp

* 20961 21716: contig of 756 bp in length

* 21717 21816: gap of 100 bp

* 21817 22579: contig of 763 bp in length

* 22580 22679: gap of 100 bp

* 22680 23440: contig of 761 bp in length

* 23441 23540: gap of 100 bp

* 23541 24305: contig of 769 bp in length

* 24310 24409: gap of 100 bp

* 24410 25142: contig of 733 bp in length

* 25143 25242: gap of 100 bp

* 25243 25963: contig of 721 bp in length

* 25964 26063: gap of 100 bp

* 26064 26770: contig of 707 bp in length

* 26771 26870: gap of 100 bp

* 26871 27628: contig of 758 bp in length

* 27629 27728: gap of 100 bp

* 27729 28488: contig of 760 bp in length

* 28489 28588: gap of 100 bp

* 28589 29333: contig of 745 bp in length

* 29334 29433: gap of 100 bp

* 29434 30183: contig of 750 bp in length

* 30184 30283: gap of 100 bp

* 30284 31026: contig of 743 bp in length

* 31027 31126: gap of 100 bp

* 31127 31888: contig of 762 bp in length

* 31889 31988: gap of 100 bp

* 31989 32737: contig of 749 bp in length

* 32738 32837: gap of 100 bp

* 32838 33575: contig of 738 bp in length

33576 33675: gap of 100 bp
33676 34405: contig of 730 bp in length
34406 34505: gap of 100 bp
34506 35232: contig of 727 bp in length
35233 35332: gap of 100 bp
35333 36044: contig of 712 bp in length
36045 36144: gap of 100 bp
36145 36871: contig of 727 bp in length
36872 36971: gap of 100 bp
36972 37677: contig of 706 bp in length
37678 37777: gap of 100 bp
37778 38514: contig of 737 bp in length
38515 38614: gap of 100 bp
38615 39351: contig of 737 bp in length
39352 39451: gap of 100 bp
39452 40205: contig of 754 bp in length
40206 40305: gap of 100 bp
40306 41069: contig of 764 bp in length
41070 41169: gap of 100 bp
41170 41890: contig of 721 bp in length
41891 41990: gap of 100 bp
41991 42718: contig of 728 bp in length
42719 42818: gap of 100 bp
42819 43542: contig of 724 bp in length
43543 43642: gap of 100 bp
43643 44393: contig of 751 bp in length
44394 44493: gap of 100 bp
44494 45236: contig of 743 bp in length
45237 45336: gap of 100 bp
45337 46080: contig of 744 bp in length
46081 46180: gap of 100 bp
46181 46927: contig of 747 bp in length
46928 47027: gap of 100 bp
47028 47725: contig of 698 bp in length
47726 47825: gap of 100 bp
47826 48570: contig of 745 bp in length
48571 48670: gap of 100 bp
48671 49435: contig of 765 bp in length
49436 49535: gap of 100 bp
49536 50280: contig of 745 bp in length
50281 50380: gap of 100 bp
50381 51105: contig of 725 bp in length
51106 51205: gap of 100 bp
51206 51946: contig of 741 bp in length
51947 52046: gap of 100 bp
52047 52778: contig of 732 bp in length
52779 52878: gap of 100 bp
52879 53636: contig of 758 bp in length
53637 53736: gap of 100 bp
53737 54475: contig of 739 bp in length
54476 54575: gap of 100 bp
54576 55332: contig of 757 bp in length
55333 55432: gap of 100 bp
55433 56200: contig of 768 bp in length
56201 56300: gap of 100 bp
56301 57071: contig of 771 bp in length
57072 57171: gap of 100 bp
57172 57926: contig of 755 bp in length
57927 58026: gap of 100 bp

Query Match 100.0%; Score 15; DB 2; Length 64671;
Best local similarity 100.0%; Pred. No. 2.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGGAATGAG 15
|||||

Db 25291 AGTATGGGGAATGAG 25305

RESULT 6

AF254984/c

LOCUS

DEFINITION Homo sapiens chromosome 8 clone RP4-611F18, WORKING DRAFT SEQUENCE, 71701 bp DNA linear HTG 24-OCT-2001

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

20 unordered pieces.
AF254984
AF254984.4 GI:16356859
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 71701)
Wen, G., Baumgart, C., Blechschmidt, K., Dettie, M., Jahn, N.,
Lehmann, R., Menzel, U., Polley, A., Reichwald, K., Schillabel, M. B.,
Schudy, A., Siddiqui, R., Taudien, S., Rosenthal, A. and Platzer, M.
Chromosome 8 genomic sequence
Unpublished
2 (bases 1 to 71701)
Polley, A., Wen, G., Baumgart, C., Dettie, M., Jahn, N., Schillabel, M.,
Menzel, U. and Rosenthal, A.
Direct Submission
Submitted (12-APR-2000) Genome Analysis, Institute of Molecular
Biotechnology, Heutenbergstrasse 11, Jena 07745, Germany
On Oct 24, 2001 this sequence version replaced qi:14327871.

Center: Genome Center
Center: Institute of Molecular Biotechnology
Center code: IMB
Web site: http://genome.imb-jena.de/
Contact: gscj-submit@genome.imb-jena.de
Project Information
Center project name: H266
Center clone name: RP4-611F18
Summary Statistics
Sequencing vector: M13; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 50971 bases at least Q40
Consensus quality: 57915 bases at least Q30
Consensus quality: 63233 bases at least Q20
Quality coverage: 3.29 x in Q20 bases; sum-of-contigs

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality 10.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 12326: contig of 12326 bp in length
12327 12426: gap of unknown length
12427 18661: contig of 6235 bp in length
18662 18761: gap of unknown length
18762 24820: contig of 5059 bp in length
24821 24920: gap of unknown length
24921 30358: contig of 5438 bp in length
30359 30458: gap of unknown length
30459 34839: contig of 4381 bp in length
34840 34939: gap of unknown length
34940 38106: contig of 3167 bp in length
38107 38206: gap of unknown length
38207 39544: contig of 1338 bp in length
39545 39644: gap of unknown length
39645 41125: contig of 1481 bp in length
41126 41225: gap of unknown length

* 41226 42665: contig of 1440 bp in length
 * 42666 42765: gap of unknown length
 * 42766 44130: contig of 1365 bp in length
 * 44131 44230: gap of unknown length
 * 44231 45564: contig of 1334 bp in length
 * 45565 45664: gap of unknown length
 * 45665 46781: contig of 1117 bp in length
 * 46782 46881: gap of unknown length
 * 46882 47973: contig of 1092 bp in length
 * 47974 48073: gap of unknown length
 * 48074 49184: contig of 1111 bp in length
 * 49185 49284: gap of unknown length
 * 49285 51393: contig of 2109 bp in length
 * 51394 51493: gap of unknown length
 * 51494 53715: contig of 2222 bp in length
 * 53716 53815: gap of unknown length
 * 53816 56375: contig of 2560 bp in length
 * 56376 56475: gap of unknown length
 * 56476 59070: contig of 2595 bp in length
 * 59071 59170: gap of unknown length
 * 59171 62293: contig of 3123 bp in length
 * 62294 62393: gap of unknown length
 * 62394 71701: contig of 9308 bp in length.

FEATURES

Location/Qualifiers
 1. .71701
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="8"
 /clone="RP4-611F18"
 BASE COUNT 20510 a 14519 c 13793 g 20979 t 1900 others
 ORIGIN

Query Match 100.0%; Score 15; DB 2; Length 71701;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGGGATGAG 15
 |||||
 Db 67056 AGTATGGGGGATGAG 67042

RESULT 7
 AL591489
 LOCUS Human DNA sequence from clone RP11-814C6 on chromosome Xq22.2-2.3,
 complete sequence.
 DEFINITION
 AL591489
 ACCESSION
 VERSION
 KEYWORDS HTG.
 SOURCE human.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 75140)
 Bray-Allen.S.
 Direct Submission
 TITLE Submitted (26-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 JOURNAL requests: clonerequests@sanger.ac.uk

COMMENT
 On Sep 27, 2001 this sequence version replaced gi:15722152.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C-elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome X, constructed by the Sanger Centre Chromosome X Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/ChrX
 RP11-814C6 is from the library RP11-11.3 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBac3.6

IMPORTANT: This sequence is not the entire insert of clone
 RP11-814C6. It may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true left end of clone RP11-733H21 is at 73141 in this
 sequence. The true right end of clone RP6-141H5 is at 2000 in this
 sequence.

FEATURES

Location/Qualifiers
 1. .75140
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="q22.2-23"
 /clone="RP11-814C6"
 /clone_lib="RP11-11.3"
 misc_feature 30621..31044
 /note="CpG island"
 /evidence=not_experimental
 BASE COUNT 23180 a 14512 c 14097 g 23351 t
 ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 75140;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGGGATGAG 15
 |||||
 Db 20336 AGTATGGGGGATGAG 20350

RESULT 8
 AC090280
 LOCUS Homo sapiens chromosome 8 clone RP4-611F18 map 8, WORKING DRAFT
 DEFINITION
 AC090280
 ACCESSION
 VERSION
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 86554)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 8, clone RP4-611F18
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 86554)
 Birren,B., Linton,L., Nusbaum,C., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Boguslavsky,L., Boukhaiter,R., Brown,A.,
 Camarata,J., Campopiano,A., Choepel,Y., Collangelo,M., Collins,S.,
 Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galadon,J.,
 Gardyna,S., Ginde,S., Goyette,M., Graham,I., Grand-pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Karatas,A., Larcocque,K., Lamazares,R., Landers,T.,
 Lehotzky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
 Marquis,N., Matthews,C., McCarney,M., McEwan,P., McKernan,K.,
 McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Milanga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,

Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Sougniez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Trivis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.

TITLE JOURNAL

COMMENT

Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 6, 2001 this sequence version replaced gi:12957916.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L12502
Center Clone name: 611_F18

----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 830% bases at least Q40
Consensus quality: 84782 bases at least Q30
Consensus quality: 85389 bases at least Q20

Insert size: 8500; agarose-1p
Insert size: 8584; sum-of-contigs
Quality coverage: 6.7 in Q20 bases; agarose-1p

Quality coverage: 6.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3966: contig of 3966 bp in length
* 3967 4066: gap of 100 bp
* 4067 8136: contig of 4070 bp in length
* 8137 8236: gap of 100 bp
* 8237 12124: contig of 3888 bp in length
* 12125 12224: gap of 100 bp
* 12225 18749: contig of 6525 bp in length
* 18750 18849: gap of 100 bp
* 18850 31371: contig of 12522 bp in length
* 31372 31471: gap of 100 bp
* 31472 43391: contig of 11920 bp in length
* 43392 43491: gap of 100 bp
* 43492 71800: contig of 28309 bp in length
* 71801 71900: gap of 100 bp
* 71901 86554: contig of 14654 bp in length.
*
* Location/Qualifiers
* 1 .86554
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="8"
* /map="8"
* /clone="RP4-611F18"
* /clone_lib="RPC1 Human PAC library 4"
* 1 .3966
* /note="assembly_fragment"
* 4067 .8136
* /note="assembly_fragment"
* 8237 .12124
* /note="assembly_fragment"
* 12225 .18749
* /note="assembly_fragment"

FEATURES source

misc_feature
misc_feature
misc_feature
misc_feature

misc_feature
misc_feature
misc_feature
misc_feature

BASE COUNT 25686 a 17956 c 17557 g 24632 t 723 others
ORIGIN

Query Match 100.0%; Score 15; DB 2; Length 86554;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
|||||
Db 45694 AGTATGGGAATGAG 45708

RESULT 9

AB018117/c

LOCUS AB018117

DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, pl clone:M015.

ACCESSION AB018117 RA000015

VERSION AB018117.1 GI:3702735

KEYWORDS

SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:M015

ORGANISM Arabidopsis thaliana

REFERENCE 1 (sites)

AUTHORS Sato, S., Nakamura, Y., Kaneko, T., Kato, T., Asamizu, E., Kotani, H. and Tabata, S.

TITLE Structural analysis of Arabidopsis thaliana chromosome 5. X.

JOURNAL DNA Res. 7 (1), 31-63 (2000)

MEDLINE 20181125

REFERENCE 2 (bases 1 to 88398)

AUTHORS Nakamura, Y.

TITLE Direct Submission

JOURNAL Submitted (06-OCT-1998) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamura@kazusa.or.jp, Tel: 81-438-52-3335, Fax: 81-438-52-3334)

COMMENT Address for correspondence: kaos@kazusa.or.jp

For the latest information on annotation of this clone, please see

http://www.kazusa.or.jp/kaos/cgi-bin/agd-graph.cgi?c=M015

Genes with similarity to proteins in the databases are described in

'product' or 'note' qualifiers. Genes that have no significant

protein similarity are described as 'unknown protein'.

The software programs used to predict genes include: Graif

(informatics group, Oak Ridge National Laboratory,

http://compbio.cornell.gov/Graif-1.3/),

GENSCAN (Chris Burge, MIT, http://CCRG-081.mit.edu/GENSCAN.html),

NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of

Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and

SplicePredictor (Volker Brendel, Stanford University,

http://gremmlini.zool.iastate.edu/cgi-bin/sp.cgi).

Genes encoding tRNAs are predicted by tRNAscan-se

(Sean Eddy, Washington University School of Medicine, St. Louis,

http://genome.wustl.edu/eddy/tRNAscan-SE/).

This sequence may not be the entire insert of this clone. It may be

shorter because we remove overlaps between neighboring submissions.

The 5' clone is K14A3 and the 3' clone is MNJ7.

Location/Qualifiers

1 .88398

/organism="Arabidopsis thaliana"

```

/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MQL5"
/clone_lib="Mitsui p1"
5073..6320
/feature="contains similarity to SET-domain protein"
gene_id:MQL5.1
/codon_start=1
/evidence=not_experimental
/protein_id="BAA97149.1"
/db_xref="GI:8809598"
/translation="MMWQRIQSPNKKRRVSFVDFPQFSVKDESIGDDVATIKN
LDGKDSNCGVAYROHHPKEESFDSIMKAGFNWANGNLGNKGFPPSKNPLPCE
QKQPLSVSEGIKLMAYESQRCFCFPLSTVKQVHRYSPAKKKLSNATLVRHS
PMKLSNARLANARHPQHDKERRSGVLSIQNRNLSKDLTPQKQVEVLRITLV
FDELDRKAARGGSETAKSRIDYQITWILREMGMVNSOKRIGSPVGIKVGDIQPK
AALSIVGLHFGIMSGIDYWKGNKEVATSVSEGNIDGDFINDVMYVCGQGMRS
KDKAIKDQKLYGNLALANSIKETPVVRVIRERRLDNRGKDYVDGLIRVEKIWE
RPGQNLDFKFKLRQCPRVDF"
complement(6850..7302)
/feature="gbIAAF26469.1"
gene_id:MQL5.2
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA97150.1"
/db_xref="GI:8809599"
/translation="MEKRSDETSSEILSDWHLSPVNPFRGIVMVSVPSPESDARL
NPPKEITRYSDKILGSTVAGEAARVLPWMREPSYEMGGKWKDGKKKKEKE
KEEIPKEIIEALLNSGDYKEDKDKYSQSCVELLKSNGFPPS"
complement(join(9726..9857,10001..10069,10500..10631,
10734..10849,10935..11016,11124..11194,11547..11607))
/feature="gene_id:MQL5.3"
/codon_start=1
/evidence=not_experimental
/product="VAMP (vesicle-associated membrane
protein)" associated protein-like"
/protein_id="BAA97151.1"
/db_xref="GI:8809600"
/translation="MTGVGNQLISIOPELKFLEKQSYCDILKVNKNTYVAK
VKTTPSKYFVPNTGVIQWDSCLIRVTLOQREYPPDMCKDKFLQSTIVPHPD
VDLPQDTFTKDSKTLTECKLKVSYITPSTQSSSGATNGDGSSETISTORLAK
PERDAAYKQQLQHELETVRRRRNRQNSGNSLKLAMVGLIGLIGLIGLITLKLAS
PT"
complement(join(12774..12882,13478..13611,13700..13786,
13958..14488))
/feature="contains similarity to plastid ribosomal protein
L19"
gene_id:MQL5.4"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA97152.1"
/db_xref="GI:8809601"
/translation="MAHKLKYTEISDKALYSGFERGLCKHDELSSSSSSSLS
HAQSHPTVASLNSVFIYHLNQALHMTPTPSPSKNGLVSYSLPRASSVNSLVS
RYFLNHSNPFPAIDSKRKEFIKABESTEGEATVENAVETAGEGEATVAEE
EAKPWKTRKGLGIMGLNKAIEAVETVPVPLGTGDIWEIKLEPKNRKLSTY
KGIWMSQNAIGHITIRLIRIAGIGVEIVPPIYSPNIKEIKVSHRKVRRLYYLR
DKLPLRSTFK"
join(16229..16282,16379..16450,16538..16693,16788..16891,
16967..17060)
/feature="gene_id:MQL5.5"
/codon_start=1
/evidence=not_experimental
/product="ras-related small GTP-binding protein-like"
/protein_id="BAA97153.1"
/db_xref="GI:8809602"
/translation="MOKIETVQDQKTKLQIWDTAQGERFTTSSYYRGAHGIYV
YVDTLSEFNWKNWLEIDRYASENVNKLNGKNDLTSQKVSFETAKAFADLGI
PFLETSKATNINVEEAMATAIKTRNASQAPGAKPPTVQIRGQPVNQSGCCSS"
complement(join(17564..17743,17828..17875,17961..18119,
18220..18361,18444..18509,18605..18923,19195..19354))

```

```

/feature="gene_id:MQL5.6"
pir||G7144
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA97154.1"
/db_xref="GI:8809603"
/translation="MASLNPFDLLGDADPSQALVALSQVFKAAAVQVPPKAKFP
TKPAPPQAVRESNAPQGGSGTGGFSGRGNGGYNRDNRNDAGNHNFGSG
YRPSDDADGNSRGSGVGYRGGRGREGPRGGVANGESDVRPRNRYHNSVGHG
TGMKNGGNGNMTTDDIPTTSEPTTEKSPVAKGGEDTPPAKRLIAERK
AKKEAEAEAREMTLEFKLLEKKALQATPKYKRVYTVFVSNQJLNNKANTDE
EFTIKLSDKDEKKDKATEKAKKSLSLINEFLPADKPKYNGRGSGRGGRGGRG
NORAKAAAPAIQDTAQFPLSG"
complement(20365..21096)
/feature="gene_id:MQL5.7"
sp|080338
/codon_start=1
/evidence=not_experimental
/product="ethylene responsive element binding factor 2
(ATERF2)"
/protein_id="BAA97155.1"
/db_xref="GI:8809604"
/translation="MYGCNIESDYALLESTRHLGGGGENELKLNSTSSSTPES
WGLPLKENDSEMLYGLLKDAFHDTSSDLSCLDFDPVAVRPTENFTAMFERPK
KAIPVTETAVKAKHYRGVRQRPWGFAAEIRDPKAKGARVIACTFTETADLAAYDIA
AFMRGSRALLNPLRVNSGEPDVRITSKRSSSSSSSSSSSTSSSENCKLKRKKAF
NLTSVQVQKCEVGDFTVDELLVS"
join(24289..24922,23278..24173)
/feature="contains similarity to non-LTR retroelement reverse
transcriptase"
gene_id:MQL5.8"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA97156.1"
/db_xref="GI:8809605"
/translation="MLEQTFSLCNLEHIVAYVPIPTDPPRPKECIATQRIYFRH
IGRDKPCGGADEETINHLIFECPPARQWALSGIPSSSRPLSSLYNLYVYWR
ANEIGACEESLRYPWIMYIWKARNKNFESICVQPOUTLDAIHEVEYWRARRE
EPOFTSPLEGGHIDMASPICFIDGSHWITDSRSHGWIILTRGRLHIGIKSRRC
LSPLHAEALWALKLDVLSICFELVKVTKDCLLTVMVNTPEMPPIFASELKUFYF
KNQLYSNIMHVPTNIRADYLAKCARTGPFYSFVSSTVILNLSNESAYP"
join(28100..28167,28395..29316)
/feature="gene_id:MQL5.9"
sp|080341"
/codon_start=1
/evidence=not_experimental
/product="ethylene responsive element binding factor 5
(ATERF5)"
/protein_id="BAA97157.1"
/db_xref="GI:8809606"
/translation="MSTIDESVIRNSGSESTKATKCFKLTLSMATPNFVSALWPIEK
HLDEASPVATDPMKHESSATSSSSSIIFGSSSSSFAPIDFSVCKPKIIDL
DTPSRMELSTPREFDESVSDDFKPSNONQRPPELKQIIRKPKIKLSIAKTE
WIOFAENTPEVTKPVSEEEKHYRGVRQRPKFAEIRHUPNKGSKWVLTFFDTA
LEARAIDEARFKSKAILNPLFVKGWKPRADECEKARKDDDEKVIYVEKVLAF
EOSVDVNGGTFFPVFVNTLTCMDWLTGFLNPLSPSPHPKGYSLQVW"
join(32469..32648,32727..32792,32884..33024,33122..33250,
33351..33494,33582..33809)
/feature="gene_id:MQL5.10"
/codon_start=1
/evidence=not_experimental
/product="mut domain protein-like"
/protein_id="BAA97158.1"
/db_xref="GI:8809607"
/translation="MNGVNLKSRITLMSAVKERSLIDAYDEWGVIVHUKLIPSNPY
AFASMLRASLSDWRKRGKVKLIPVEQSELVPIAKGCFYHHAEGYVNLITWIP
EEESFPMANASHQVGVGVGVVQKVELVYVQEKYCAPSITGLKLPCTGFINESIEIF
SCAVRYKEGTVDTEFSEVIAFRHNAVAFESDLFFICMLPLSLKIIIDALEIKA
AKWPLFAEFVQPMIRGDKMRKRVTEICEARLSHRVCGLSPHRIVSTFDGKPPSLVYN
VDDDDHPDSHNSCTEFYR"
complement(34387..36918)
/feature="gene_id:MQL5.11"

```

```

/codon_start=1
/evidence=not_experimental
/product="NBS/LRR disease resistance protein"
/protein_id="BA097159.1"
/db_xref="GI:8809608"

Query Match      100.0%; Score 15; DB 8; Length 88398;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
|||||
Db 76572 AGTATGGGAATGAG 76558

RESULT 11
AL513353
LOCUS      105225 bp      DNA      linear      PRI 26-SEP-2001
DEFINITION Human DNA sequence from clone Rp11-309N12 on chromosome 9, complete
sequence.
ACCESSION  AL513353 AC041045
VERSION     AL513353.7 GI:15795469
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 105225)
AUTHORS     Corby, N.
TITLE       Direct Submission
JOURNAL     Submitted (26-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 27, 2001 this sequence version replaced gi:14702141.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
Rp11-309N12 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
Rp11-309N12 it may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true right end of clone Rp11-309N12 is at 105225 in this
sequence. The true left end of clone Rp11-446A5 is at 11971 in this
sequence. The true right end of clone Rp11-305L7 is at 2000 in this
sequence.

FEATURES             Location/Qualifiers
     source           1..105225
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="Rp11-309N12"

```

```

/clone_lib="RPCI-11.2"
BASE COUNT      34150 a 21427 c 19042 g 30606 t
ORIGIN
Query Match      100.0%; Score 15; DB 9; Length 105225;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
|||||
Db 2080 AGTATGGGAATGAG 2094

RESULT 11
AL136374/c
LOCUS      119853 bp      DNA      linear      PRI 18-JUL-2000
DEFINITION Human DNA sequence from clone Rp1-244G5 on chromosome 1q24.3 25.3,
complete sequence.
ACCESSION  AL136374
VERSION     AL136374.4 GI:8919204
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 119853)
AUTHORS     Kimberley, A.
TITLE       Direct Submission
JOURNAL     Submitted (17-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 5, 2000 this sequence version replaced gi:18176662.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
Rp1-244G5 is from the library RPCI-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone Rp1-244G5.

FEATURES             Location/Qualifiers
     source           1..119853
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="q24.3-25.3"
                     /clone="Rp1-244G5"
                     /clone_lib="RPCI-1"
     repeat_region    1..237
                     /note="AluX repeat: matches 1..237 of consensus"
     repeat_region    240..925
                     /note="L2 repeat: matches 1995..2706 of consensus"
                     complement(1507..1926)
     misc_feature      /note="match: GSS: Em:AQ146487"
     misc_feature      1822..2311

```

```

repeat_region /note="match: GSS: Em:AQ624491"
1830..1956
repeat_region /note="L2 repeat: matches 2616..2746 of consensus"
2127..2181
repeat_region /note="L2 repeat: matches 2656..2710 of consensus"
2380..2691
repeat_region /note="AluY repeat: matches 1..311 of consensus"
2716..3083
misc_feature /note="MLTIB repeat: matches 1..390 of consensus"
complement(2776..3233)
/note="match: STS: Em:G37725
3252..3690
misc_feature /note="match: GSS: Em:AQ773761"
4104..4288
repeat_region /note="MIR repeat: matches 8..219 of consensus"
4533..4756
repeat_region /note="L2 repeat: matches 1992..2228 of consensus"
4915..5144
repeat_region /note="MIR repeat: matches 2..252 of consensus"
5284..5483
repeat_region /note="L2 repeat: matches 2541..2749 of consensus"
7056..7103
repeat_region /note="L2 copies 4 mer tcca 85% conserved"
7216..7391
repeat_region /note="MIR repeat: matches 73..262 of consensus"
8076..8276
repeat_region /note="MIR repeat: matches 61..262 of consensus"
8458..8494
repeat_region /note="MIR repeat: matches 31..67 of consensus"
13489..13811
repeat_region /note="L2 repeat: matches 1301..1648 of consensus"
13837..13892
repeat_region /note="L4 copies 4 mer gtgt 75% conserved"
14392..15214
repeat_region /note="MER45B repeat: matches 1..835 of consensus"
16024..16224
repeat_region /note="MIR repeat: matches 34..248 of consensus"
16428..16737
repeat_region /note="AluY repeat: matches 1..299 of consensus"
complement(17098..17761)
/note="match: GSS: Em:AQ477207"
17714..18068
repeat_region /note="L2 repeat: matches 2155..2537 of consensus"
18073..18710
repeat_region /note="L2 repeat: matches 1049..1745 of consensus"
complement(18237..18836)
/note="match: GSS: Em:AQ527362"
complement(18332..19056)
/note="match: GSS: Em:AQ309729"
complement(18876..19056)
/note="match: STS: Em:HSPE58B10"
18880..19060
repeat_region /note="match: STS: Em:HSPE23B05"
20092..20163
repeat_region /note="L2 repeat: matches 2673..2746 of consensus"
20198..20725
repeat_region /note="L2 repeat: matches 189..691 of consensus"
20726..20992
repeat_region /note="AluSc repeat: matches 1..299 of consensus"
20993..21176
repeat_region /note="L2 repeat: matches 1..189 of consensus"
21234..21896
repeat_region /note="L2 repeat: matches 1..660 of consensus"
22078..22430
repeat_region /note="L2 repeat: matches 3..385 of consensus"
22651..23063
repeat_region /note="MLTIC repeat: matches 50..466 of consensus"
23449..23783
repeat_region /note="MER2 repeat: matches 1..345 of consensus"
complement(24518..25322)
/note="match: GSS: Em:AQ745887"
24785..24962

```

```

repeat_region /note="MIR repeat: matches 49..262 of consensus"
30110..30149
repeat_region /note="20 copies 2 mer gg 77% conserved"
30419..30718
misc_feature /note="AluY repeat: matches 1..299 of consensus"
30754..31346
repeat_region /note="match: GSS: Em:AQ392308"
30834..31024
repeat_region /note="MER58A repeat: matches 26..207 of consensus"
31661..31780
repeat_region /note="MIR repeat: matches 2..115 of consensus"
31784..32115
repeat_region /note="L1PA5 repeat: matches 5960..6294 of consensus"
32517..32610
repeat_region /note="MIR repeat: matches 100..194 of consensus"
32626..32939
repeat_region /note="AluSp repeat: matches 1..313 of consensus"
33275..33494
repeat_region /note="L1PA5 repeat: matches 5922..6145 of consensus"
33620..33844
repeat_region /note="MIR repeat: matches 20..261 of consensus"
33976..34701
repeat_region /note="L2 repeat: matches 1..691 of consensus"
35234..35696
repeat_region /note="MLTIC repeat: matches 1..466 of consensus"
35981..36020
repeat_region /note="10 copies 4 mer gaga 92% conserved"
36393..36458
repeat_region /note="L2 repeat: matches 2682..2746 of consensus"
36546..36711
repeat_region /note="MER5B repeat: matches 2..173 of consensus"
complement(37350..37907)
/note="match: GSS: Em:AQ801709"
37578..37810
repeat_region /note="AluSg/X repeat: matches 77..299 of consensus"
37844..37947
repeat_region /note="L2 repeat: matches 2589..2710 of consensus"
37938..38503
repeat_region /note="match: GSS: Em:AQ727965"
complement(38576..39060)
/note="match: GSS: Em:AQ721135"
39270..39399
repeat_region /note="MIR repeat: matches 90..220 of consensus"
40557..41199
repeat_region /note="L2 repeat: matches 1882..2529 of consensus"
41533..42009
repeat_region /note="match: GSS: Em:AQ403526"
42947..43072
repeat_region /note="MER5A repeat: matches 64..189 of consensus"
43714..43818
repeat_region /note="L2 repeat: matches 2374..2478 of consensus"
44053..44085
repeat_region /note="L2 repeat: matches 2673..2705 of consensus"
44219..44274
repeat_region /note="28 copies 2 mer tt 78% conserved"
44275..44569
repeat_region /note="L1PA5 repeat: matches 5849..6143 of consensus"
45083..45140
repeat_region /note="AluSp/q repeat: matches 181..243 of consensus"
45141..45400
repeat_region /note="AluSc repeat: matches 1..265 of consensus"
45402..45515
repeat_region /note="MER5A repeat: matches 51..187 of consensus"
46114..46312
repeat_region /note="L2 repeat: matches 5962..6154 of consensus"
46321..47484
repeat_region /note="L2 repeat: matches 4307..5529 of consensus"
47489..48076
repeat_region /note="L1M4 repeat: matches 3246..3849 of consensus"

```

Query Match 100.0%; Score 15; DB 9; Length 119853;
 Best Local Similarity 100.0%; Pred. NO. 2.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
 |||||
 Db 27722 AGTATGGGAATGAG 27708

RESULT 12
 AC003659

LOCUS
 DEFINITION Homo sapiens clone RPI-146A15, WORKING DRAFT SEQUENCE, 61 unordered pieces.
 AC003659
 VERSION AC003659.5 GI:11128350
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 137400)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Boucek,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bulay,C., Burci,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cart,N.T.F., Carter,M., Cavaros,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,A., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okwona,G., Oragunye,N., Owiedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shocshtari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,X., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., and Gibbs,R.

TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 137400)
 AUTHORS Werley,K.C.

TITLE Direct Submission
 JOURNAL Submitted (02-DEC-1997) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 COMMENT On Nov 9, 2000 this sequence version replaced gi:9739307.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information

 Center project name: UH

Center clone name: RPI-146A15
 ----- Summary Statistics
 Sequencing vector: M13; L08821
 Chemistry: Dye-primer Bodipy: 9% of reads
 Chemistry: Dye-terminator Big Dye: 7% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 105729 bases at least Q40
 Consensus quality: 120136 bases at least Q30
 Consensus quality: 128905 bases at least Q20
 Estimated insert size: 104322; sum-of-contrigs estimation
 Estimated insert size: 160000; agarose-fp estimation
 Quality coverage: 2.7x in Q20 bases; agarose-fp estimation
 Quality coverage: 4.2x in Q20 bases; sum-of-contrigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 61 contrigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contrigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 17502: contig of 17502 bp in length
 * 17503 17602: gap of unknown length
 * 17603 22207: contig of 4605 bp in length
 * 22208 22307: gap of unknown length
 * 22308 26935: contig of 4628 bp in length
 * 26936 30407: gap of unknown length
 * 30408 30507: gap of unknown length
 * 30508 35891: contig of 5384 bp in length
 * 35892 35991: gap of unknown length
 * 35992 38909: contig of 2918 bp in length
 * 38910 39009: gap of unknown length
 * 39010 42540: contig of 3531 bp in length
 * 42541 42640: gap of unknown length
 * 42641 44851: contig of 2211 bp in length
 * 44852 48257: contig of 3306 bp in length
 * 48258 48357: gap of unknown length
 * 48358 51101: contig of 2744 bp in length
 * 51102 51201: gap of unknown length
 * 51202 53932: contig of 2691 bp in length
 * 53933 53993: gap of unknown length
 * 53994 56279: contig of 2286 bp in length
 * 56280 56378: gap of unknown length
 * 56379 59081: contig of 2703 bp in length
 * 59082 59181: gap of unknown length
 * 59182 61855: contig of 2674 bp in length
 * 61856 61955: gap of unknown length
 * 61956 63676: contig of 1721 bp in length
 * 63677 63776: gap of unknown length
 * 63777 66425: contig of 2849 bp in length
 * 66426 66525: gap of unknown length
 * 66526 69062: contig of 2537 bp in length
 * 69063 69162: gap of unknown length
 * 69163 71681: contig of 2519 bp in length
 * 71682 71781: gap of unknown length
 * 71782 73793: contig of 2012 bp in length
 * 73794 73893: gap of unknown length
 * 73894 76681: contig of 2788 bp in length
 * 76682 78419: gap of unknown length
 * 78420 78519: contig of 1638 bp in length
 * 78520 79713: gap of unknown length
 * 79714 79813: contig of 1194 bp in length
 * 79814 81543: gap of unknown length
 * 81544 81643: contig of 1730 bp in length
 * 81644 83090: gap of unknown length
 * 83091 83190: contig of 1447 bp in length
 * 83191 84733: contig of 1543 bp in length


```

* 84734 84833: gap of unknown length
* 84834 86577: contig of 1744 bp in length
* 86578 86577: gap of unknown length
* 86578 86578: contig of 1596 bp in length
* 88273 88273: gap of unknown length
* 88374 90502: contig of 2129 bp in length
* 90503 90502: gap of unknown length
* 90603 91783: contig of 1181 bp in length
* 91784 91883: gap of unknown length
* 91884 93498: contig of 1615 bp in length
* 93499 93498: gap of unknown length
* 93599 94947: contig of 1349 bp in length
* 94948 95047: gap of unknown length
* 95048 96477: contig of 1430 bp in length
* 96478 96577: gap of unknown length
* 96578 98103: contig of 1526 bp in length
* 98104 98203: gap of unknown length
* 98204 99264: contig of 1061 bp in length
* 99265 100833: gap of unknown length
* 99365 100833: contig of 1469 bp in length
* 100834 100933: gap of unknown length
* 100934 101965: contig of 1032 bp in length
* 101966 102065: gap of unknown length
* 102066 103994: contig of 1929 bp in length
* 103995 104094: gap of unknown length
* 104095 106329: contig of 2235 bp in length
* 106330 106429: gap of unknown length
* 106430 107485: contig of 1056 bp in length
* 107486 107585: gap of unknown length
* 107586 109189: contig of 1604 bp in length
* 109190 109289: gap of unknown length
* 109290 110951: contig of 1862 bp in length
* 110952 111051: gap of unknown length
* 111053 112053: contig of 1002 bp in length
* 112054 112153: gap of unknown length
* 112154 113185: contig of 1032 bp in length
* 113186 113285: gap of unknown length
* 113286 114879: contig of 1594 bp in length
* 114880 114979: gap of unknown length
* 114980 116140: contig of 1161 bp in length
* 116141 116240: gap of unknown length
* 116241 117460: contig of 1220 bp in length
* 117461 117560: gap of unknown length
* 117561 118914: contig of 1354 bp in length
* 118915 119014: gap of unknown length
* 119015 120326: contig of 1312 bp in length
* 120327 120426: gap of unknown length
* 120427 121655: contig of 1229 bp in length
* 121656 121755: gap of unknown length
* 121756 122853: contig of 1098 bp in length
* 122854 122953: gap of unknown length
* 122954 124447: contig of 1494 bp in length
* 124448 124547: gap of unknown length
* 124548 126289: contig of 1742 bp in length
* 126290 126389: gap of unknown length
* 126390 127545: contig of 1156 bp in length
* 127546 127645: gap of unknown length
* 127646 128733: contig of 1088 bp in length
* 128734 128833: gap of unknown length
* 128834 130104: contig of 1271 bp in length
* 130105 130204: gap of unknown length
* 130205 131309: contig of 1105 bp in length

```

Query Match 100.0%; Score 15; DB 2; Length 137400;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGTATGGGGAATGAG 15
 |||||
 Db 3011 AGTATGGGGAATGAG 3025

RESULT 13

AC004147
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

TITLE
 JOURNAL

REFERENCE
 AUTHORS

TITLE
 JOURNAL

REFERENCE
 AUTHORS

TITLE
 JOURNAL

COMMENT

AC004147 139049 bp DNA linear PRI 06-JUN-2000
 Homo sapiens chromosome 17, clone C11304117, complete sequence.
 AC004147
 AC004147.1 GI:2995605
 HTG.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 139049)
 Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P. and
 Lander,E.

Unpublished

2 (bases 1 to 139049)

Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P.,
 Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Becklerly,R.,
 Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cerny,J., Cooke,P.,
 Daly,M.J., Depayre,E., Devon,K., Dewar,K., Donelan,L., DuRette,B.,
 Etemadi,S., Ferreira,P., Forrest,C., Funke,R., Gage,D., Gaidyana,S.,
 Gensheimer,S., Geraldery,K., Gilmartin,T., Gray,D., Hagos,B.,
 Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Linton,L.,
 Mackenzie,J., Marquis,N., McEwan,P., McGurk,A., Meldrum,J.,
 Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J., O'Connor,T.,
 Pavlin,B., Peterson,K., Ranganath,S., Riley,R., Roberts,D.,
 Rollins,G., Rossello,R., Roy,A., Shyam,R., Soohoo,S.,
 Stange-Thomann,N., Stillwell,J., Stone,C., Strickland,C., S. Huey,K.,
 Tang,J., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y.,
 Ye,W.J., Zemtseva,I., Zhao,J. and Zody,M.

Direct Submission

Submitted (14-FEB-1998) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 139049)

Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P.,
 Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Becklerly,R.,
 Bann,J., Boatn,C., Boutwell,C., Brown,A., Byrne,S., Cantu,C.,
 Castle,A., Cerny,J., Cooke,P., Daly,M.J., Depayre,E., Devon,K.,
 Dewar,K., Donelan,L., DuRette,B., Etemadi,S., Ferreira,P.,
 FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gaidyana,S.,
 Gensheimer,S., Geraldery,K., Gilmartin,T., Grant,G., Gray,D.,
 Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L.,
 Kann,L., Linton,L., Macdonald,P., Marquis,N., McEwan,P., McGurk,A.,
 Meldrum,J., Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J.,
 Naylor,J., O'Connor,T., Pavlin,B., Peterson,K., Riley,R.,
 Roberts,D., Rollins,G., Rossello,R., Roy,A., Shyam,R.,
 Stange-Thomann,N., Stillwell,J., Stone,C., Strickland,C.,
 Subramanian,A., Sydney,K., Tang,L., Vassiliev,H., Vo,A., Wagner,A.,
 Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.

Direct Submission

Submitted (28-MAR-1998) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 139049)

Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P.,
 Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Becklerly,R.,
 Bann,J., Boatn,C., Boutwell,C., Brown,A., Byrne,S., Cantu,C.,
 Castle,A., Cerny,J., Cooke,P., Daly,M.J., Depayre,E., Devon,K.,
 Dewar,K., Donelan,L., DuRette,B., Etemadi,S., Ferreira,P.,
 FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gaidyana,S.,
 Gensheimer,S., Geraldery,K., Gilmartin,T., Grant,G., Gray,D.,
 Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L.,
 Kann,L., Linton,L., Macdonald,P., Marquis,N., McEwan,P., McGurk,A.,
 Meldrum,J., Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J.,
 Naylor,J., O'Connor,T., Pavlin,B., Peterson,K., Riley,R.,
 Roberts,D., Rollins,G., Rossello,R., Roy,A., Shyam,R.,
 Stange-Thomann,N., Stillwell,J., Stone,C., Strickland,C.,
 Subramanian,A., Sydney,K., Tang,L., Vassiliev,H., Vo,A., Wagner,A.,
 Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.

Direct Submission

Submitted (06-JUN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 28, 1998 this sequence version replaced gi:2967619.

All repeats were identified using RepeatMasker: Smit, A.F.A. &

Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

FEATURES

Source

```

1. 139049
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="GIT-304117"
/clone_lib="ClifC Human BAC"
complement(12..208)
/rpt_family="MER30"
416..439
/rpt_family="AT-rich"
complement(441..738)
/rpt_family="AluJo"
1290..1413
/rpt_family="POLY-A"
complement(2984..3179)
/rpt_family="MIR"
3903..4055
/rpt_family="MIR"
4206..4318
/rpt_family="MIR"
4324..4542
/rpt_family="MER20"
complement(5646..5803)
/rpt_family="MIR"
5909..5985
/rpt_family="MIR"
complement(7086..7142)
8419..8719
/rpt_family="(TAAAA)n"
/rpt_family="AluJo"
complement(9161..9372)
/rpt_family="LINE2"
9539..9619
/rpt_family="MERSA"
11072..11327
/rpt_family="MIR"
complement(11597..11743)
/rpt_family="MIR"
complement(12392..12424)
/rpt_family="(CA)n"
12493..12724
/rpt_family="MIR"
13356..13480
/rpt_family="(CA)n"
complement(15748..15780)
/rpt_family="(CA)n"
18631..18719
/rpt_family="MIR"
complement(19314..19340)
/rpt_family="AT-rich"
19383..19529
/rpt_family="MIR"
complement(20229..20513)
/rpt_family="AluSg"
complement(20804..21671)
/rpt_family="L1MA8"
22031..22246
/rpt_family="AluJb"
22247..22279
/rpt_family="(GAAAA)n"
22302..22324
/rpt_family="AT-rich"
24057..24121
/rpt_family="MIR"
complement(24283..24316)
/rpt_family="(TAAA)n"
24692..24977
/rpt_family="LTR16A"
26655..26784

```

```

repeat_region      /rpt_family="MIR"
complement(27085..27296)
repeat_region      /rpt_family="MIR"
complement(27385..27537)
repeat_region      /rpt_family="MSTC"
complement(27752..28040)
repeat_region      /rpt_family="MSTC"
complement(28043..28164)
repeat_region      /rpt_family="MIR"
complement(29708..29846)
repeat_region      /rpt_family="LINE2"
complement(29966..30050)
repeat_region      /rpt_family="L1MC4"
complement(30389..31942)
repeat_region      /rpt_family="L1MC4"
complement(33029..33197)
repeat_region      /rpt_family="MIR"
33376..33733
/rpt_family="MLT1A1"
complement(34560..35178)
repeat_region      /rpt_family="MER4D"
complement(35177..35407)
repeat_region      /rpt_family="MER4D"
35547..35578
/rpt_family="AT-rich"
complement(35749..35909)
repeat_region      /rpt_family="MERSB"
complement(37344..37553)
repeat_region      /rpt_family="MER20"
38656..39153
/rpt_family="LTR12"
complement(39154..39253)
repeat_region      /rpt_family="MIR"
complement(40534..40644)
repeat_region      /rpt_family="LINE2"
40801..40883
/rpt_family="LINE2"
42080..42224
/rpt_family="MIR"
42430..42528
/rpt_family="MERSB"
complement(42774..42839)
repeat_region      /rpt_family="MIR"
complement(43354..43413)
repeat_region      /rpt_family="MIR"
44462..44552
/rpt_family="MERSA"
45602..45680
/rpt_family="MIR"
46417..46803
/rpt_family="MLT1A2"
complement(47106..47146)
repeat_region      /rpt_family="MIR"
47153..47250
/rpt_family="MIR"

```

Query Match

Best Local Similarity 100.0%; score 15; DB 9; Length 139049;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACTATGGGGAATGAG 15

DB 26942 ACTATGGGGAATGAG 26956

RESULT 14

AC023037
LOCUS AC023037 153787 bp DNA linear HTG 01-MAR-2000
DEFINITION Homo sapiens chromosome X clone RP11-630D8 map X, WORKING DRAFT
ACCESSION AC023037
VERSION AC023037.2 GI:7139835
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

GenCore version 5.1.3

Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 08:45:30 ; Search time 145 Seconds
(without alignments)
177.612 Million cell updates/sec

Title: US-09-820-203A-1

Perfect score: 15

Sequence: 1 agtatgggaatgag 15

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	15	100.0	1354	21 AAC46782	Arabidopsis thaliana
C 2	15	100.0	6156	19 AA03517	Human transcriptio
C 3	14	93.3	246	21 AAF14719	Aspergillus oryzae
C 4	14	93.3	422	22 AAS3524	Human cDNA encodin
C 5	14	93.3	554	22 AAH1458	Human cDNA clone (
C 6	14	93.3	832	21 AAF13418	Aspergillus oryzae
C 7	14	93.3	2798	22 AAH16206	Human cDNA sequenc
C 8	13.4	89.3	461	22 ABA58130	Human foetal liver
C 9	13.4	89.3	461	22 ABA27334	Probe #5800 for ge

10	13.4	89.3	461	22 AAC06209	Human brain expres
11	13.4	89.3	461	22 AAC31868	Human bone marrow
12	13.4	89.3	461	22 AA115850	Probe #5783 for ge
13	13.4	89.3	461	22 AA137736	Probe #6422 used t
C 14	13.4	89.3	511	22 AA091494	Human digestive sy
15	13.4	89.3	517	22 ABA60977	Human foetal liver
16	13.4	89.3	517	22 ABA28922	Human brain expres
17	13.4	89.3	517	22 AAK09269	Human brain expres
18	13.4	89.3	517	22 AAK35158	Human bone marrow
19	13.4	89.3	517	22 AAK140874	Probe #9560 used t
20	13.4	89.3	537	22 AAS42492	Human cDNA encodin
21	13.4	89.3	575	22 ABA63120	Human foetal liver
22	13.4	89.3	575	22 ABA30371	Probe #8837 for ge
23	13.4	89.3	575	22 AAK11548	Human brain expres
24	13.4	89.3	575	22 AAK37320	Human bone marrow
25	13.4	89.3	575	22 AAK18151	Probe #8084 for ge
26	13.4	89.3	575	22 AA143159	Probe #11845 used
27	13.4	89.3	576	23 AAS52814	Enterococcus faeca
C 28	13.4	89.3	738	20 AAX27431	Human secreted pro
29	13.4	89.3	819	22 AAH03929	Human cDNA clone (
30	13.4	89.3	955	22 AAS34046	Human cDNA encodin
C 31	13.4	89.3	1332	21 AAC79685	Human secreted pro
32	13.4	89.3	1344	21 AAC77685	Human cancer assoc
33	13.4	89.3	1398	13 AAQ20254	Encodes beta-chain
34	13.4	89.3	1398	13 AAQ22142	RBC1 cDNA coding f
35	13.4	89.3	1398	22 AAH44782	Hepaticocyte growth
C 36	13.4	89.3	1446	22 ABA89359	Escherichia coli p
37	13.4	89.3	1843	22 AAH15386	Human cDNA sequenc
38	13.4	89.3	1873	20 AAZ33672	Human breast tumou
39	13.4	89.3	1874	20 AAZ42121	Human endometrium
40	13.4	89.3	2121	23 AAS64230	DNA encoding novel
C 41	13.4	89.3	2187	13 AAQ20255	Encodes alpha- and
C 42	13.4	89.3	2187	21 AAZ29701	Wild-type human c-
C 43	13.4	89.3	2187	22 AAH28358	Nucleotide sequenc
C 44	13.4	89.3	2222	22 AA199327	Human excretory re
45	13.4	89.3	2222	22 AA163677	Human kidney relat

ALIGNMENTS

RESULT 1
AAC46782/c
ID AAC46782 standard; DNA: 1354 BP.
AC AAC46782;
XX
XX
DT 18-OCT-2000 (first entry)
XX
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 51383.
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139849.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-APR-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147250.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151055.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159337.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 15; DB 21; Length 1354;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
 ID 894 AGTATGGGAATGAG 880
 DB 894 AGTATGGGAATGAG 880

RESULT 2
 AAV03517/C
 ID AAV03517 standard; cDNA; 6156 BP.
 XX AAV03517;
 AC AAV03517;
 DT 20-JUL-1998 (first entry)
 XX Human transcriptional intermediary factor-2 (TIP2) cDNA.
 DE Transcriptional intermediary factor: TIP2; human; drug screening;
 XX assay; nuclear receptor; ds.
 KW Homo sapiens.
 OS
 GX
 FH Key Location/Qualifiers
 FT CDS 163..4557
 FT /tag= a
 XX
 PN W09802455-A2.
 XX
 XX 22-JAN-1998.
 XX 11-JUL-1997; 97WO-US12100.
 XX 12-JUL-1996; 96US-0021247.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 PA (CNRS) CENT NAT RECH SCI.
 PA (INEM) INST NAT SANTE & RECH MEDICALE.
 PA (UYPA) UNIV PASTEUR LOUIS.
 XX Chambon P, Gronemeyer H, Lutz Y, Voegel J;
 XX WPI: 1998-110525/10.
 DR P-PSDB; AAW42632.
 XX
 PT New isolated transcriptional intermediary factor-2 - useful in
 assays for drugs which are capable of enhancing or inhibiting

PT nuclear receptor-mediated pathways
 XX Claim 1; Fig 1A-B; 119pp; English.
 PS
 XX This cDNA clone codes for a novel nuclear receptor transcriptional
 CC mediator (see AAW42632) designated transcriptional intermediary
 CC factor-2 (TIF2). TIF2 interacts directly with the ligand binding
 CC domains (LBD) of several nuclear receptors (NR) in an agonist- and
 CC AF2-integrity-dependent manner in vitro and in vivo, harbours an
 CC autonomous AF, relieves NR autoquenching, and enhances the
 CC activity of NR AFs when overexpressed in mammalian cells.
 CC Screening of a human placenta cDNA expression library with an
 CC estradiol-bound 32P-labelled estrogen receptor probe yielded a cDNA
 CC that encoded a protein fragment (TIF2.1) that interacted on
 CC Far-Western blots with different 32P-labelled NR-LBDs ER, RAR, RXR
 CC in an agonist-dependent manner. The TIF2 coding sequence was
 CC obtained upon rescuing with a TIF2.1 cDNA probe. TIF2 cDNA has
 CC been deposited as ATCC 97612. Polynucleotides encoding full-length
 CC TIF2 protein or encoding TIF2 polypeptide fragments comprising
 CC amino acids 624-869, 624-1131, 1010-1131 or 1288-1464 are claimed.
 CC Recombinant methods for making TIF2 polypeptides are provided, as
 CC are screening methods for identifying agonists and antagonists of
 CC nuclear receptor AF-2 function, TIF2 AD1 activity and TIF2 AD2
 CC activity. The products are useful in assays for identifying drugs
 CC capable of enhancing or inhibiting NR-mediated pathways. They can
 CC also be used for detection and localisation.
 XX Sequence 6156 BP; 1799 A; 1490 G; 1406 C; 1461 T; 0 other;
 SQ

Query Match 100.0%; Score 15; DB 19; Length 6156;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
 ID 5420 AGTATGGGAATGAG 5406
 DB 5420 AGTATGGGAATGAG 5406

RESULT 3
 AAF14719/C
 ID AAF14719 standard; cDNA; 246 BP.
 XX AAF14719;
 AC AAF14719;
 DT 13-MAR-2001 (first entry)
 XX Aspergillus oryzae EST SEQ ID NO:7242.
 DE
 XX Multiple gene expression: filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Aspergillus oryzae.
 XX
 XX WO2000056762-A2.
 XX 28-SEP-2000.
 XX 22-MAR-2000; 2000WO-US07781.
 XX 22-MAR-1999; 99US-0273623.
 PR (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen RH;
 XX WPI: 2000-594572/56.
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a


```

PT substrate of expressed sequence tags -
PS Claim 88; Page 2933; 3161pp; English.
XX
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (ff) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC ff cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (ff) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from ff cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how ff cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 246 BP; 72 A; 67 C; 56 G; 51 T; 0 other;

Query Match 93.3%; Score 14; DB 21; Length 246;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGTATGGGAATGA 14
   |||||
Db 51 AGTATGGGAATGA 38

RESULT 4
ID AAS33524
AC AAS33524;
XX
XX 17-DEC-2001 (first entry)
XX
{ Human cDNA encoding a novel foetal antigen, SEQ ID No 48.
KW Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;
KW immunomodulator; cardiovascular; cytostatic; nephrothropic;
KW cardiovascular; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; breast neoplasm; cancer;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; food additive.
XX
OS Homo sapiens.
XX
XX WO200155312-A2.
PN
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01321.
XX
XX 31-JAN-2000; 2000US-0179065.
PR
PR 04-FEB-2000; 2000US-0180628.
PR
PR 24-FEB-2000; 2000US-0184664.
PR
PR 02-MAR-2000; 2000US-0186350.
PR
PR 16-MAR-2000; 2000US-0189874.
PR
PR 17-MAR-2000; 2000US-0190076.
PR

```

```

PR 18-APR-2000; 2000US-0198123.
PR
PR 19-MAY-2000; 2000US-0205515.
PR
PR 07-JUN-2000; 2000US-0209467.
PR
PR 28-JUN-2000; 2000US-0214886.
PR
PR 30-JUN-2000; 2000US-0215135.
PR
PR 07-JUL-2000; 2000US-0216647.
PR
PR 07-JUL-2000; 2000US-0216880.
PR
PR 11-JUL-2000; 2000US-0217487.
PR
PR 11-JUL-2000; 2000US-0217496.
PR
PR 14-JUL-2000; 2000US-0218290.
PR
PR 26-JUL-2000; 2000US-0220963.
PR
PR 26-JUL-2000; 2000US-0220964.
PR
PR 14-AUG-2000; 2000US-0224518.
PR
PR 14-AUG-2000; 2000US-0224519.
PR
PR 14-AUG-2000; 2000US-0225213.
PR
PR 14-AUG-2000; 2000US-0225214.
PR
PR 14-AUG-2000; 2000US-0225266.
PR
PR 14-AUG-2000; 2000US-0225267.
PR
PR 14-AUG-2000; 2000US-0225268.
PR
PR 14-AUG-2000; 2000US-0225270.
PR
PR 14-AUG-2000; 2000US-0225447.
PR
PR 14-AUG-2000; 2000US-0225757.
PR
PR 14-AUG-2000; 2000US-0225758.
PR
PR 14-AUG-2000; 2000US-0225759.
PR
PR 18-AUG-2000; 2000US-0226279.
PR
PR 22-AUG-2000; 2000US-0226681.
PR
PR 22-AUG-2000; 2000US-0226868.
PR
PR 22-AUG-2000; 2000US-0227182.
PR
PR 23-AUG-2000; 2000US-0227009.
PR
PR 30-AUG-2000; 2000US-0228924.
PR
PR 01-SEP-2000; 2000US-0229287.
PR
PR 01-SEP-2000; 2000US-0229343.
PR
PR 01-SEP-2000; 2000US-0229344.
PR
PR 01-SEP-2000; 2000US-0229345.
PR
PR 05-SEP-2000; 2000US-0229509.
PR
PR 05-SEP-2000; 2000US-0229513.
PR
PR 06-SEP-2000; 2000US-0230437.
PR
PR 06-SEP-2000; 2000US-0230438.
PR
PR 08-SEP-2000; 2000US-0231242.
PR
PR 08-SEP-2000; 2000US-0231243.
PR
PR 08-SEP-2000; 2000US-0231244.
PR
PR 08-SEP-2000; 2000US-0231413.
PR
PR 08-SEP-2000; 2000US-0231414.
PR
PR 08-SEP-2000; 2000US-0232080.
PR
PR 08-SEP-2000; 2000US-0232081.
PR
PR 12-SEP-2000; 2000US-0231968.
PR
PR 14-SEP-2000; 2000US-0232397.
PR
PR 14-SEP-2000; 2000US-0232398.
PR
PR 14-SEP-2000; 2000US-0232399.
PR
PR 14-SEP-2000; 2000US-0232400.
PR
PR 14-SEP-2000; 2000US-0232401.
PR
PR 14-SEP-2000; 2000US-0233063.
PR
PR 14-SEP-2000; 2000US-0233064.
PR
PR 21-SEP-2000; 2000US-0233065.
PR
PR 21-SEP-2000; 2000US-0234223.
PR
PR 21-SEP-2000; 2000US-0234274.
PR
PR 25-SEP-2000; 2000US-0234597.
PR
PR 25-SEP-2000; 2000US-0234598.
PR
PR 26-SEP-2000; 2000US-0234584.
PR
PR 27-SEP-2000; 2000US-0235834.
PR
PR 27-SEP-2000; 2000US-0235836.
PR
PR 29-SEP-2000; 2000US-0236327.
PR
PR 29-SEP-2000; 2000US-0236367.
PR
PR 29-SEP-2000; 2000US-0236368.
PR
PR 29-SEP-2000; 2000US-0236369.
PR
PR 29-SEP-2000; 2000US-0236370.
PR
PR 02-OCT-2000; 2000US-0236602.
PR
PR 02-OCT-2000; 2000US-0237037.
PR
PR 02-OCT-2000; 2000US-0237038.
PR
PR 02-OCT-2000; 2000US-0237039.
PR
PR 02-OCT-2000; 2000US-0237040.
PR
PR 13-OCT-2000; 2000US-0239935.
PR
PR 13-OCT-2000; 2000US-0239937.
PR

```

PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.

PR Rosen CA, Barash SC, Ruben SM;
 PI WPI: 2001-488782/53.
 DR P-PSDB; ANU20704.
 XX

PT New polynucleotides and polypeptides for diagnosing, treating,
 PT preventing or prognosing e.g. diseases or disorders of the nervous,
 PT musculoskeletal, excretory, gastrointestinal, reproductive, and
 PT respiratory systems -

XX Claim 1; SEQ ID No 48; 642pp; English.

XX The invention relates to novel nucleic acids encoding novel human foetal

CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
 CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. The antibodies to the antigens can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. Numerous
 CC examples of diseases and disorders treated by the nucleic acids and
 CC proteins are given in the specification. The present sequence

Query Match 93.3%; Score 14; DB 22; Length 422;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACTATGGGGAATGA 14

|||||||

Db 347 AGTATGGGGAATGA 360

RESULT 5

AAH11458/C

ID AAH11458 standard; cDNA; 554 BP.

XX AC AAH11458;

XX DT 26-JUN-2001 (first entry)

XX DE Human cDNA clone (3'-primer) SEQ ID NO:8293.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -

XX Claim 3; SEQ ID 8293; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 554 BP; 215 A; 92 C; 99 G; 144 T; 4 other;

Query Match 93.3%; Score 14; DB 22; Length 554;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGTATGGGAATGA 14
 | | | | | | | | | |
 Db 32 AGTATGGGAATGA 19

RESULT 6

AAH13418/C
 ID AAF13418 standard; cDNA; 832 BP.

XX AC AAF13418;

XX DT 13-MAR-2001 (first entry)

XX DE Aspergillus oryzae EST SEQ ID NO:5941.

XX Multiple gene expression; filamentous fungal cell; EST:
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.

OS Aspergillus oryzae.

XX PN WO2000056762-A2.

XX PD 28-SEP-2000.

XX PF 22-MAR-2000; 2000WO-US07781.

XX PR 22-MAR-1999; 99US-0273623.

XX (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.

XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 PI WPI; 2000-594572/56.

XX DR Monitoring differential expression of genes in filamentous fungal cells
 PI uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -

XX Claim 88; Page 2453-2454; 3161pp; English.

CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organization of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from
 CC Aspergillus niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.

XX Sequence 832 BP; 230 A; 224 C; 182 G; 196 T; 0 other;

Query Match 93.3%; Score 14; DB 21; Length 832;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGTATGGGAATGA 14

| | | | | | | | | |

Db 94 AGTATGGGAATGA 81

RESULT 7

AAH16206
 ID AAH16206 standard; cDNA; 2798 BP.

XX AC AAH16206;

XX DT 26-JUN-2001 (first entry)

XX DE Human cDNA sequence SEQ ID NO:15011.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -

XX PS Claim 8; SEQ ID 15011; 2537pp + CD ROM; English.

XX CC The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to

CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX S0 Sequence 2798 BP; 716 A; 627 C; 678 G; 777 T; 0 other;

Query Match 93.3%; Score 14; DB 22; Length 2798;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGTATGGGGAATGA 14

Db 2767 AGTATGGGGAATGA 2780

|||||

RESULT 8

ABA58130

ID ABA58130 standard; DNA; 461 BP.

XX AC ABA58130;

XX DT 01-FEB-2002 (first entry)

XX DE Human foetal liver single exon nucleic acid probe #6435.

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX JS Homo sapiens.

XX PN WO200157277-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00669.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX DR Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human fetal liver.

XX PS Claim 1; SEQ ID NO 6435; 639pp + sequence listing; English.

XX CC The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human

CC fetal liver. The present sequence is a single exon nucleic acid

CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX S0 Sequence 461 BP; 150 A; 60 C; 109 G; 142 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 461;

Best Local Similarity 93.3%; Pred. No. 4.2e+02;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AGTATGGGGAATGAG 15

Db 168 AGTGTGGGGAATGAG 182

|||||

RESULT 9

ABA27334

ID ABA27334 standard; DNA; 461 BP.

XX AC ABA27334;

XX DT 23-JAN-2002 (first entry)

XX DE Probe #5800 for gene expression analysis in human heart cell sample.

XX KW Human; gene expression; heart; microarray; vascular system; probe;

XX KW cardiovascular disease; hypertension; cardiac arrhythmia;

XX KW congenital heart disease; ss.

XX OS Homo sapiens.

XX PN WO200157274-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00666.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX DR Single exon nucleic acid probes for analyzing gene expression in human

XX PT hearts -

XX PS Claim 1; SEQ ID No 5800; 530pp; English.

XX CC The present invention relates to single exon nucleic acid probes for

CC measuring human gene expression in a sample derived from human heart. The

CC present sequence is one such probe. The probes may be used for

CC predicting, measuring and displaying gene expression in samples derived

CC from the human heart via microarrays. By measuring gene expression, the

CC probes are useful for predicting, diagnosing, grading, staging,

CC monitoring and prognosing diseases of the human heart and vascular system

CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and

CC congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 461 BP; 150 A; 60 C; 109 G; 142 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 461;
 Best Local Similarity 93.3%; Pred. No. 4.2e+02;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
 ||| |||||

DB 168 AGTGTGGGAATGAG 182

RESULT 10
 AAK06209
 ID AAK06209 standard; DNA; 461 BP.

AAK06209;

05-NOV-2001 (first entry)

Human brain expressed single exon probe SEQ ID NO: 6200.

Human; brain expressed exon; gene expression analysis; probe:
 microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 epilepsy; cancer; ss.

Homo sapiens.

WO200157275-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00667.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-483446/52.

Single exon nucleic acid probes for analyzing gene expression in human
 brains -

Example 4; SEQ ID NO: 6200; 650pp + Sequence Listing; English.

The present invention provides a number of single exon nucleic acid
 probes which are derived from genomic sequences expressed in the human
 brain. They can be used to measure gene expression in brain cell samples,
 which may enable the diagnosis and improved treatment of nervous system
 diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 epilepsy and cancers. The present sequence is one of the probes of the
 invention.

Sequence 461 BP; 150 A; 60 C; 109 G; 142 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 461;
 Best Local Similarity 93.3%; Pred. No. 4.2e+02;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
 ||| |||||

DB 168 AGTGTGGGAATGAG 182

RESULT 11

AAK31868
 ID AAK31868 standard; DNA; 461 BP.

AAK31868;

06-NOV-2001 (first entry)

Human bone marrow expressed single exon probe SEQ ID NO: 6425.

Human; bone marrow expressed exon; gene expression analysis; probe:
 microarray; cancer; leukaemia; lymphoma; myeloma; ss.

Homo sapiens.

WO200157276-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00668.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488900/53.

Human genome-derived single exon nucleic acid probes useful for
 analyzing gene expression in human bone marrow -

Example 4; SEQ ID NO: 6425; 658pp + Sequence Listing; English.

The present invention provides a number of single exon nucleic acid
 probes which are derived from genomic sequences expressed in the human
 bone marrow. They can be used to measure gene expression in bone marrow
 samples, which may enable the improved diagnosis and treatment of cancers
 such as lymphoma, leukaemia and myeloma. The present sequence is one of
 the probes of the invention.

Sequence 461 BP; 150 A; 60 C; 109 G; 142 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 461;

Best Local Similarity 93.3%; Pred. No. 4.2e+02;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
 ||| |||||

DB 168 AGTGTGGGAATGAG 182

RESULT 12

AAI15850
 ID AAI15850 standard; DNA; 461 BP.

AAI15850;

12-OCT-2001 (first entry)

Probe #5783 for gene expression analysis in human cervical cell samples.
 Probe; human; microarray; gene expression; cervical epithelial cell;
 cervical cancer; ss.

```

XX OS Homo sapiens.
XX PN WO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00670.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488901/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human cervical epithelial cells.
XX PS Claim 25; SEQ ID No 5783; 487pp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENPs). The present sequence is one such probe. The SENPs are derived
XX CC from human HeLa cells. The SENPs can be used to produce a single exon
XX CC microarray, which can be used for measuring human gene expression in a
XX CC sample derived from human cervical epithelial cells. By measuring gene
XX CC expression, the probes are therefore useful in grading and/or staging
XX CC of diseases of the cervix, notably cervical cancer.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SO Sequence 461 BP; 150 A; 60 C; 109 G; 142 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 461;
Best Local Similarity 93.3%; Pred. No. 4.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTATGGGGAATGAG 15
   ||| ||| ||| ||| |||
Db 168 AGTGTGGGGAATGAG 182

RESULT 13
AAI37736
ID AAI37736 standard; DNA; 461 BP.
XX AC AAI37736;
XX DT 17-OCT-2001 (first entry)
XX DE Probe #6422 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; es.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.

```

```

PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-48897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta.
XX PS Claim 25; SEQ ID No 6422; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENPs).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX SO Sequence 461 BP; 150 A; 60 C; 109 G; 142 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 461;
Best Local Similarity 93.3%; Pred. No. 4.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTATGGGGAATGAG 15
   ||| ||| ||| ||| |||
Db 168 AGTGTGGGGAATGAG 182

RESULT 14
AAK91494/c
ID AAK91494 standard; DNA; 511 BP.
XX AC AAK91494;
XX DT 05-NOV-2001 (first entry)
XX DE Human digestive system antigen genomic sequence SEQ ID NO: 5070.
XX KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
XX KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
XX KW digestive system disorder; Meckel's diverticulum; ds.
XX OS Homo sapiens.
XX PN WO200155314-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01324.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 14-JUL-2000; 2000US-0217496.
XX PR 26-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.

```

PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250150.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-502630/55.
XX
XX Polynucleotides encoding digestive system antigens, useful for
PI diagnosing, treating, preventing and/or prognosing disorders of the
PI digestive system, particularly cancer and cancer metastases.
XX
XX Disclosure; SEQ ID NO 5070; 986pp; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
XX encoding a digestive system antigen of the invention.
SQ Sequence 511 BP; 120 A; 103 C; 92 G; 196 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 511;

Best Local Similarity 93.3%; Pred. No. 4.2e-02;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
 || |||||
 Db 69 AGAATGGGAATGAG 55

RESULT 15

ABA60977
 ID ABA60977 standard; DNA: 517 BP.
 AC ABA60977;
 XX 01-FEB-2002 (first entry)
 DT Human foetal liver single exon nucleic acid probe #9282.
 DE Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 KW Homo sapiens.
 XX WO200157277-A2.
 OS 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00669.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI: 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX Claim 1: SEQ ID NO 9282; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 C fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 517 BP; 98 A; 136 C; 173 G; 110 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 517;

Best Local Similarity 93.3%; Pred. No. 4.2e+02;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15

Db 294 AGGATGGGAATGAG 308

Search completed: October 17, 2002, 10:09:03
 Job time : 149 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 08:51:00 ; Search time 32 Seconds
(without alignments)
115.141 Million cell updates/sec

Title: US-09-820-203A-1
Perfect score: 15
Sequence: 1 agtatgggaatgag 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 38533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCIVS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
C 1	15	100.0	6156	4	US-08-891-640-1	Sequence 1, Appl
C 2	15	100.0	6156	4	US-09-723-535-3	Sequence 3, Appl
C 3	13.4	89.3	738	4	US-09-227-357-131	Sequence 131, App
C 4	13.4	89.3	4038	3	US-08-969-125-8	Sequence 8, Appl
5	13.4	89.3	9751	4	US-09-238-303-7	Sequence 7, Appl
6	13.4	89.3	152331	3	US-09-128-155-16	Sequence 16, Appl
C 7	13.4	89.3	176373	3	US-09-128-155-17	Sequence 17, Appl
8	13	86.7	4181	1	US-08-072-574-9	Sequence 9, Appl
9	13	86.7	4181	1	US-08-486-270-9	Sequence 9, Appl
10	13	86.7	4181	3	US-08-367-264-9	Sequence 9, Appl
11	13	86.7	4303	4	US-08-660-148-4	Sequence 4, Appl
12	13	86.7	4303	4	US-08-660-148-6	Sequence 6, Appl
13	12.4	82.7	24	3	US-09-085-735-3	Sequence 3, Appl
14	12.4	82.7	24	3	US-09-085-442-3	Sequence 3, Appl
15	12.4	82.7	676	4	US-08-687-3808-5	Sequence 5, Appl
16	12.4	82.7	1313	2	US-08-463-911-6	Sequence 1, Appl
C 17	12.4	82.7	1365	1	US-08-114-072-1	Sequence 1, Appl
C 18	12.4	82.7	1365	5	PCT-US94-09361-1	Sequence 34, Appl
C 19	12.4	82.7	1664	1	US-08-250-740-34	Sequence 3, Appl
C 20	12.4	82.7	1664	1	US-07-695-472B-3	Sequence 102, App
C 21	12.4	82.7	2005	4	US-08-936-165A-102	Sequence 9, Appl
22	12.4	82.7	2162	3	US-09-013-881-9	Sequence 35, Appl
C 23	12.4	82.7	2344	1	US-07-695-472B-35	Sequence 32, Appl
C 24	12.4	82.7	2347	1	US-08-250-740-32	Sequence 1, Appl
C 25	12.4	82.7	2347	1	US-07-695-472B-34	Sequence 34, Appl
C 26	12.4	82.7	2347	1	US-07-695-472B-34	Sequence 36, Appl
C 27	12.4	82.7	2347	1	US-07-695-472B-36	

C 28 12.4 82.7 2469 3 US-09-087-727-1 Sequence 1, Appl
29 12.4 82.7 4517 4 US-09-140-804-9 Sequence 9, Appl
C 30 12.4 82.7 4741 1 US-07-695-472B-4 Sequence 4, Appl
C 31 12.4 82.7 4742 1 US-08-250-740-35 Sequence 35, Appl
C 32 12.4 82.7 5436 2 US-08-948-277A-2 Sequence 2, Appl
C 33 12.4 82.7 5436 2 US-09-169-203-2 Sequence 3, Appl
C 34 12.4 82.7 59065 4 US-09-813-817-3 Sequence 20, Appl
C 35 12.4 82.7 246240 2 US-08-724-394A-20 Sequence 21, Appl
C 36 12.4 82.7 246240 2 US-08-724-394A-21 Sequence 22, Appl
C 37 12.4 82.7 246240 2 US-08-724-394A-22 Sequence 11, Appl
38 12 80.0 27 1 US-08-482-282B-11 Sequence 11, Appl
39 12 80.0 27 1 US-08-486-036A-11 Sequence 18, Appl
40 12 80.0 27 5 PCT-US96-09848-11 Sequence 9, Appl
41 12 80.0 27 1 US-08-482-282B-9 Sequence 9, Appl
42 12 80.0 35 1 US-08-486-036A-9 Sequence 9, Appl
43 12 80.0 35 5 PCT-US96-09848-9 Sequence 69, Appl
44 12 80.0 340 4 US-08-836-075A-69

ALIGNMENTS

RESULT 1
US-08-891-640-1/c
; Sequence 1, Application US/08891640
; Patent No. 6268173
; GENERAL INFORMATION:
; APPLICANT: Chamboon, Pierre
; APPLICANT: Gronemeyer, Hinrich
; APPLICANT: Voegel, Johannes
; APPLICANT: Lutz, Yves
; TITLE OF INVENTION: Transcriptional Intermediary Factor-2
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0. Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,640
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/021,247
; FILING DATE: 12-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1383.0130001/EKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6156 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 163..4554
US-08-891-640-1

Query Match 100.0%; Score 15; DB 4; Length 6156;

Best Local Similarity 100.0%; Pred. No. 13;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 1 AGTATGGGAATGAG 15

Db 5420 AGTATGGGAATGAG 5406

RESULT 2

US-09-723-535-3/c
; Sequence 3, Application US/09723535
; Patent No. 6355483
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Lex M. Cowsett
; TITLE OF INVENTION: ANTISENSE MODULATION OF SRC-2 EXPRESSION
; FILE REFERENCE: RTS-0225
; CURRENT APPLICATION NUMBER: US/09/723.535
; CURRENT FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 49
; SEQ ID NO 3
; LENGTH: 6156
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (163)...(4557)
US-09-723-535-3

Query Match 100.0%; Score 15; DB 4; Length 6156;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 1 AGTATGGGAATGAG 15

Db 5420 AGTATGGGAATGAG 5406

RESULT 3

US-09-227-357-131/c
; Sequence 131, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010p1
; CURRENT APPLICATION NUMBER: US/09/227.357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920

; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-227-357-131

Query Match 89.3%; Score 13.4; DB 4; Length 738;
Best Local Similarity 93.3%; Pred. No. 78;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Caps 0;

Qy 1 AGTATGGGAATGAG 15

Db 116 AGTATGGGAATGAG 102

RESULT 4

US-08-969-125-8
; Sequence 8, Application US/08969125B
; Patent No. 6143871
; GENERAL INFORMATION:
; APPLICANT: BONNEFOY, JEAN-YVES
; GAUCHAT, JEAN-FRANCOIS
; TITLE OF INVENTION: SUBSTANCES AND THEIR USES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.

ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/969,125B
FILING DATE: 12-NO. 6143871-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9625899.1
FILING DATE: 13-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: WILSON, MARY J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 1430-179
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 8
SEQUENCE CHARACTERISTICS:
LENGTH: 4038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FRAGMENT TYPE: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 43..1323
SEQUENCE DESCRIPTION: SEQ ID NO: 8
US-08-969-125-8

Query Match 89.3%; Score 13.4; DB 3; Length 4038;
Best Local Similarity 93.3%; Pred. No. 89;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTATGGGGAATGAG 15
|||||

DB 3047 AGTATGGGGAATGAG 3061

RESULT 5
US-09-238-303-7
; Sequence 7, Application US/09238303B
; Patent No. 6284253
; APPLICANT: Barr, Margaret C.
; TITLE OF INVENTION: No. 6284253el Feline Immunodeficiency Virus Nucleotide Sequence
; FILE REFERENCE: 18617, 0059
; CURRENT APPLICATION NUMBER: US/09/238,303B
; CURRENT FILING DATE: 1999-01-28
; EARLIER APPLICATION NUMBER: US 60/072,927
; EARLIER FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 7
; LENGTH: 9751
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: recombinant viral clone constructed from the genomic DNA of
; OTHER INFORMATION: a Pallas's cat feline immunodeficiency virus
US-09-238-303-7

Query Match 89.3%; Score 13.4; DB 4; Length 9751;
Best Local Similarity 93.3%; Pred. No. 96;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTATGGGGAATGAG 15
|||||

DB 682 AGGATGGGGAATGAG 696

RESULT 6
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 89.3%; Score 13.4; DB 3; Length 152331;
Best Local Similarity 93.3%; Pred. No. 11e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTATGGGGAATGAG 15
|||||

DB 19396 ATTATGGGGAATGAG 19410

RESULT 7
US-09-128-155-17/c
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

Query Match 89.3%; Score 13.4; DB 3; Length 176373;
Best Local Similarity 93.3%; Pred. No. 11e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTATGGGGAATGAG 15
|||||

DB 123833 ATTATGGGGAATGAG 123819

RESULT 8

US-08-072-574-9
; Sequence 9, Application US/08072574
; Patent No. 5521297
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; TITLE OF INVENTION: HUMAN MATABOTROPIC GLUTAMATE RECEPTORS.
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 00719

```
Query Match      86.7%; Score 13; DB 1; Length 4181;
Best Local Similarity 100.0%; pred. No. 1.5e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 AGTATGGGGAATG 13
 |||||
 Db 3061 AGTATGGGGAATG 3073

RESULT 9

RES-001, 9
 US-08-486-270-9
 ; Sequence 9, Application US/08486270
 ; Patent No. 5807689
 ; GENERAL INFORMATION:
 ; APPLICANT: Daggett, Lorrie
 ; APPLICANT: Ellis, Steven B.
 ; APPLICANT: Liaw, Chen
 ; APPLICANT: Pontsler, Aaron
 ; APPLICANT: Johnson, Edwin C.
 ; APPLICANT: Hess, Stephen D.
 ; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
 ; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
 ; NUMBER OF SEQUENCES: 13

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-BOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,270
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/072,574
FILING DATE: 04-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: FP41 9772
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4181 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..4008
OTHER INFORMATION: /product= "HUMAN MGLUR5B"
OTHER INFORMATION: /note= "variant of MGLUR5A with
OTHER INFORMATION: insertion between nucleotides 2
85-08-486-270-9

```

```
Query Match: 86.7%; Score 13; DB 1; Length 4181;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 AGTATGGGAATG 13
 |||||
 Db 3061 AGTATGGGAATG 3073

RESULT 10
US-08-367-264-9
; Sequence 9, Application US/08367264

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/367,264
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/072,574
FILING DATE: 04-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: FP41 9772
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4181 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..4008
OTHER INFORMATION: /product= "HUMAN MGLUR5B"
OTHER INFORMATION: /note= "Variant of MGLUR5A with 96 base pair
insertion between nucleotides 2998 and 2999."
US-08-367-264-9

Query Match 86.7%; Score 13; DB 3; Length 4181;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGGGAATG 13
|||||
DB 3061 AGTATGGGGGAATG 3073

RESULT 11
US-08-660-148-4
Sequence 4, Application US/08660148
Patent No. 6211353
GENERAL INFORMATION:
APPLICANT: Burnett, J. P.
APPLICANT: Mayne, Nancy G.
APPLICANT: Sharp, Robert L.
APPLICANT: Snyder, Yvonne M.
TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
RELATED NUCLEIC ACID COMPOUNDS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,148
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/282,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-9419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 4303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: mRNA
US-08-660-148-6

NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-9419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 4303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 460..4095
US-08-660-148-4

Query Match 86.7%; Score 13; DB 4; Length 4303;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGGGAATG 13
|||||
DB 3151 AGTATGGGGGAATG 3163

RESULT 12
US-08-660-148-6
Sequence 6, Application US/08660148
Patent No. 6211353
GENERAL INFORMATION:
APPLICANT: Burnett, J. P.
APPLICANT: Mayne, Nancy G.
APPLICANT: Sharp, Robert L.
APPLICANT: Snyder, Yvonne M.
TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
RELATED NUCLEIC ACID COMPOUNDS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,148
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/282,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-9419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 4303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: mRNA
US-08-660-148-6

Query Match 86.7%; Score 13; DB 4; Length 4303;
Best Local Similarity 76.9%; Pred. No. 1.5e+02;
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGAATG 13
|||||
Db 3151 AGAUGGGGAUG 3163

RESULT 13

US-09-085-735-3
; Sequence 3, Application US/09085735
; Patent No. 6072107

GENERAL INFORMATION:

APPLICANT: Latch, Garrick C
APPLICANT: Christensen, Michael J
APPLICANT: Tapper, Brian A
APPLICANT: Easton, Herrick S
APPLICANT: Hume, David E
APPLICANT: Fletcher, Lester R
TITLE OF INVENTION: Ryegrass Endophytes
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee Winner and Sullivan
STREET: 5370 Manhattan Circle #201
CITY: Boulder
STATE: CO
COUNTRY: US
ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/085,735
FILING DATE: 27-MAY-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: NZ 314926
FILING DATE: 27-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Greenlee, Lorraine L
REGISTRATION NUMBER: 27,894
REFERENCE/DOCKET NUMBER: 43-98
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303-499-8080
TELEFAX: 303-499-8080

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 24 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "oligonucleotide"
HYPOTHETICAL: NO

US-09-085-735-3

Query Match 82.7%; Score 12.4; DB 3; Length 24;
Best Local Similarity 92.9%; Pred. No. 2.1e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GTATGGGAATG 15
|||||
Db 7 GTATGGGAATG 20

RESULT 14

US-09-085-442-3
; Sequence 3, Application US/09085442
; Patent No. 611170

GENERAL INFORMATION:
APPLICANT: Latch, Garrick C
APPLICANT: Christensen, Michael J
APPLICANT: Tapper, Brian A
APPLICANT: Easton, Herrick S
APPLICANT: Hume, David E
APPLICANT: Fletcher, Lester R
TITLE OF INVENTION: Hill Rescue Endophytes
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee Winner and Sullivan
STREET: 5370 Manhattan Circle #201
CITY: Boulder
STATE: CO
COUNTRY: US
ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/085,442
FILING DATE: 27-MAY-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: NZ 314925
FILING DATE: 27-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Greenlee, Lorraine L
REGISTRATION NUMBER: 27,894
REFERENCE/DOCKET NUMBER: 44-98
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303-499-8080
TELEFAX: 303-499-8080
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "oligonucleotide"
HYPOTHETICAL: NO

US-09-085-442-3

Query Match 82.7%; Score 12.4; DB 3; Length 24;
Best Local Similarity 92.9%; Pred. No. 2.1e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GTA"GGGAATG 15
|||||
Db 7 GTATGGGAATG 20

RESULT 15

US-08-687-5808-5
; Sequence 5, Application US/086875808
; Patent No. 6291647

GENERAL INFORMATION:

APPLICANT: MELCHERS, Leo Sjoerd
APPLICANT: PONSTEIN, Anne Silene
APPLICANT: KROON-SWART, Saskia
APPLICANT: VAN DEVENTER-TROOST, Johanna Petronella Els
APPLICANT: OHL, Stephan Andres
APPLICANT: BRES-VLOEMANS, Alexandra Aleida
APPLICANT: LOGEMANN, J rgen
APPLICANT: SELA-BUURLAGE, Marianne Beatrix
TITLE OF INVENTION: Antifungal proteins, DNA coding therefor, and
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: LADAS & PARRY

```

; STREET: 26 WEST 61 STREET
; CITY: NEW YORK
; STATE: NY
; ZIP: 10023
; COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3-1/4" Disk 1.44MB
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Microsoft Windows for Workgroups 3.11
; SOFTWARE: WordPerfect 8 for Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,580B
; FILING DATE: 20-NOV-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/00488
; FILING DATE: 2-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NL94200321.1
; FILING DATE: 9-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: MASS, Clifford J.
; REGISTRATION NUMBER: 30,086
; REFERENCE/DOCKET NUMBER: U-010885-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 708 1890
; TELEFAX: (212) 246 8959
; TELEX: 233288
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 676 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Nicotiana tabacum
; STRAIN: Samsun NN
; IMMEDIATE SOURCE:
; CLONE: PCR-A
; US-08-687-580B-5

```

```

Query Match      82.7%; Score 12.4; DB 4; Length 676;
Best Local Similarity 92.9%; Pred. No. 2.7e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AGTATGGGAATGA 14
Db 611 AGTATGGGAACGA 624

```

```

Search completed: October 17, 2002, 10:11:48
Job time : 115 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 08:52:30 ; Search time 958 Seconds
(without alignments)
327.660 Million cell updates/sec

Title: US-09-820-203A-2

Perfect score: 15

Sequence: 1 acaatttccacga 15

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pt.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vt.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vt.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description

1	15	100.0	1563	1	BTYI6269	
2	15	100.0	1840	1	BTPLC	
3	15	100.0	2017	1	BACPLCA	
4	15	100.0	2254	1	BTPIPLC1	
5	15	100.0	15399	6	AX346416	
6	15	100.0	27722	3	AC006717	
7	15	100.0	38199	3	AC009745	
8	15	100.0	40018	2	AC018103	
9	15	100.0	44334	3	AF077534	
10	15	100.0	87286	8	AB026651	
11	15	100.0	94647	10	MUSTCRA	
12	15	100.0	110000	10	AE008686_3	
13	15	100.0	152948	2	AP003776	
14	15	100.0	169400	2	AC022208	
15	15	100.0	169556	9	AP003062	
16	15	100.0	175179	3	AC009254	
17	15	100.0	180911	2	AC067946	
18	15	100.0	194277	9	AC096865	
19	15	100.0	197872	2	AC037476	
20	15	100.0	220878	9	AC095050	
21	15	100.0	269619	3	CEV51H4A	
22	15	100.0	311800	3	AE003781	
23	14	93.3	159	14	ADRI7RB	
24	14	93.3	900	1	COXCEBA	
25	14	93.3	1780	8	AF265483	
26	14	93.3	2350	3	AE003396	
27	14	93.3	12580	8	SPAC1296	
28	14	93.3	26521	3	CEW01D2	
29	14	93.3	32243	3	CEV58E10	
30	14	93.3	40748	2	AC108263	
31	14	93.3	40906	3	U46668	
32	14	93.3	41754	3	AF016450	
33	14	93.3	61969	2	AC014103	
34	14	93.3	64607	9	AC005021	
35	14	93.3	72666	9	AC093614	
36	14	93.3	74100	3	AC090839	
37	14	93.3	77350	8	AT25815	
38	14	93.3	93890	8	AC006528	
39	14	93.3	100139	9	AC021752	
40	14	93.3	101463	2	AC098771	
41	14	93.3	106935	3	AC087079	
42	14	93.3	110716	2	AC008717	
43	14	93.3	110932	3	CEV38E10A	
44	14	93.3	115177	9	AL583836	
45	14	93.3	119526	9	AL559267	

ALIGNMENTS

RESULT 1
BTYI6269/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BTYI6269 1563 bp DNA linear BCT 02-01-1998
Bacillus thuringiensis pipic gene.
Y16269
Y16269.1 GI:2815229
phospholipase C; pipic gene.
Bacillus thuringiensis.
Bacillus thuringiensis
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group.
1 (bases 1 to 1563)
Lovgren,A., Carlson,C.R., Eskils,K. and Kolsto,A.B.
Localization of putative virulence genes on a physical map of the
Bacillus thuringiensis subsp. gelechiae chromosome
Curr. Microbiol. 37 (4), 245-250 (1998)
98404099
2 (bases 1 to 1563)
Lovgren,A.
Direct Submission
Submitted (21-JAN-1998) A. Lovgren, Dept of Microbiology, Stockholm
University, S-106 91 Stockholm 106 91, SWEDEN
Location/Qualifiers

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES


```
source
1..1563
/organism="Bacillus thuringiensis"
/strain="t13"
/sub_species="gelechiae"
/db_xref="taxon:1428"
/clone="pip1b"
/sub_clone="pip1.6"
767..1563
/gene="pip1c"
767..>1563
/gene="pip1c"
/codon_start=1
/transl_table=11
/product="phosphatidylinositol-specific phospholipase C"
/db_xref="GI:2815230"
/translation="MSKKLLIKLFCSTIFITFVFALHDKRVVAASSVNELENSKW
MQIPDNIPLARISIPGTHSDGTFKLQNP:KQVGMTOEYDFRYOMDHGAR:FD:RGR
LTDNDITVLHGGPLYLYVTLHEF:INEAKOFLKDNPSSETI:IMSLKKEYEDMKGAGSFS
STPEKNYFVDPIFLKTEGNIKLGDAKGI:VLLKRYSGSNEGGTNNIT:TWPNIT:FTTT
VYQNVAVTQDQKYKYNDEKYSIKDTMDTMMNSDNLNHY:INT:SI:SSGGT:AWNSP
YYIASYINPELANDIKQKNP:TRVGW:IQDY:INEKWSPLLYQEV:IRANKSI:IKP"
BASE COUNT 611 a 200 c 285 g 467 t
ORIGIN
1 ACAATTTTCCACGA 15
|||||
Db 1335 ACAATTTTCCACGA 1321

RESULT 2
BTSPC/c
LOCUS
DEFINITION
Bacillus thuringiensis DNA for phosphatidylinositol-specific
phospholipase C (EC 3.1.4.10 ).
ACCESSION
X12952
VERSION
1
KEYWORDS
1-phosphatidylinositol phosphodiesterase;
phosphatidylinositol-specific phospholipase C; phospholipase.
SOURCE
Bacillus thuringiensis.
ORGANISM
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus thuringiensis.
REFERENCE
1 (bases 1 to 1840)
Hennen,D.J.
Direct Submission
Submitted (20-SEP-1988) Hennen D.J., Genentech, Inc., Dept. of Cell
Genetics, 460 Point San Bruno Boulevard, South San Francisco, CA
94080, USA
2 (bases 601 to 1840)
Hennen,D.J., Yang,M., Chen,E., Hellmiss,R., Rodriguez,H. and
Low,M.G.
Sequence of the Bacillus thuringiensis phosphatidylinositol
specific phospholipase C
Nucleic Acids Res. 16 (21), 10383 (1988)
89057495
FEATURES
Location/Qualifiers
1..1840
/organism="Bacillus thuringiensis"
/strain="ATCC 10792"
/db_xref="taxon:1428"
761..853
761..1750
/codon_start=1
/product="prephospholipase C (AA -31 to 298)"
/protein_id="CAA31410.1"
/db_xref="GI:40328"
/db_xref="SWISS-PROT:P08954"
/translation="MSKKLLIKLFCSTIFITFVFALHDKRVVAASSVNELENSKW

sig_peptide
761..1750
CDS
1..1563
/organism="Bacillus thuringiensis"
/strain="t13"
/sub_species="gelechiae"
/db_xref="taxon:1428"
/clone="pip1b"
/sub_clone="pip1.6"
767..1563
/gene="pip1c"
767..>1563
/gene="pip1c"
/codon_start=1
/transl_table=11
/product="phosphatidylinositol-specific phospholipase C"
/db_xref="GI:2815230"
/translation="MSKKLLIKLFCSTIFITFVFALHDKRVVAASSVNELENSKW
MQIPDNIPLARISIPGTHSDGTFKLQNP:KQVGMTOEYDFRYOMDHGAR:FD:RGR
LTDNDITVLHGGPLYLYVTLHEF:INEAKOFLKDNPSSETI:IMSLKKEYEDMKGAGSFS
STPEKNYFVDPIFLKTEGNIKLGDAKGI:VLLKRYSGSNEGGTNNIT:TWPNIT:FTTT
VYQNVAVTQDQKYKYNDEKYSIKDTMDTMMNSDNLNHY:INT:SI:SSGGT:AWNSP
YYIASYINPELANDIKQKNP:TRVGW:IQDY:INEKWSPLLYQEV:IRANKSI:IKP"
BASE COUNT 611 a 200 c 285 g 467 t
ORIGIN
1 ACAATTTTCCACGA 15
|||||
Db 1335 ACAATTTTCCACGA 1321

Query Match 100.0%; Score 13; DB 1; Length 1563;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAATTTTCCACGA 15
|||||
Db 1335 ACAATTTTCCACGA 1321

RESULT 2
BTSPC/c
LOCUS
DEFINITION
Bacillus thuringiensis DNA for phosphatidylinositol-specific
phospholipase C (EC 3.1.4.10 ).
ACCESSION
X12952
VERSION
1
KEYWORDS
1-phosphatidylinositol phosphodiesterase;
phosphatidylinositol-specific phospholipase C; phospholipase.
SOURCE
Bacillus thuringiensis.
ORGANISM
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus thuringiensis.
REFERENCE
1 (bases 1 to 1840)
Hennen,D.J.
Direct Submission
Submitted (20-SEP-1988) Hennen D.J., Genentech, Inc., Dept. of Cell
Genetics, 460 Point San Bruno Boulevard, South San Francisco, CA
94080, USA
2 (bases 601 to 1840)
Hennen,D.J., Yang,M., Chen,E., Hellmiss,R., Rodriguez,H. and
Low,M.G.
Sequence of the Bacillus thuringiensis phosphatidylinositol
specific phospholipase C
Nucleic Acids Res. 16 (21), 10383 (1988)
89057495
FEATURES
Location/Qualifiers
1..1840
/organism="Bacillus thuringiensis"
/strain="ATCC 10792"
/db_xref="taxon:1428"
761..853
761..1750
/codon_start=1
/product="prephospholipase C (AA -31 to 298)"
/protein_id="CAA31410.1"
/db_xref="GI:40328"
/db_xref="SWISS-PROT:P08954"
/translation="MSKKLLIKLFCSTIFITFVFALHDKRVVAASSVNELENSKW

sig_peptide
761..1750
CDS
1..1563
/organism="Bacillus thuringiensis"
/strain="t13"
/sub_species="gelechiae"
/db_xref="taxon:1428"
/clone="pip1b"
/sub_clone="pip1.6"
767..1563
/gene="pip1c"
767..>1563
/gene="pip1c"
/codon_start=1
/transl_table=11
/product="phosphatidylinositol-specific phospholipase C"
/db_xref="GI:2815230"
/translation="MSKKLLIKLFCSTIFITFVFALHDKRVVAASSVNELENSKW
MQIPDNIPLARISIPGTHSDGTFKLQNP:KQVGMTOEYDFRYOMDHGAR:FD:RGR
LTDNDITVLHGGPLYLYVTLHEF:INEAKOFLKDNPSSETI:IMSLKKEYEDMKGAGSFS
STPEKNYFVDPIFLKTEGNIKLGDAKGI:VLLKRYSGSNEGGTNNIT:TWPNIT:FTTT
VYQNVAVTQDQKYKYNDEKYSIKDTMDTMMNSDNLNHY:INT:SI:SSGGT:AWNSP
YYIASYINPELANDIKQKNP:TRVGW:IQDY:INEKWSPLLYQEV:IRANKSI:IKP"
BASE COUNT 611 a 200 c 285 g 467 t
ORIGIN
1 ACAATTTTCCACGA 15
|||||
Db 1335 ACAATTTTCCACGA 1321

Query Match 100.0%; Score 15; DB 1; Length 1840;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAATTTTCCACGA 15
|||||
Db 1329 ACAATTTTCCACGA 1315

RESULT 3
BACPLCA/c
LOCUS
DEFINITION
B. cereus phosphatidylinositol-specific phospholipase C gene.
ACCESSION
M30809
VERSION
M30809.1
KEYWORDS
phospholipase.
SOURCE
B. cereus (ATCC 6454) DNA, clone pA2-delta55.
ORGANISM
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group.
REFERENCE
1 (bases 1 to 2017)
Kupke,A., Evans,L.M., McMillen,D.A. and Griffith,O.H.
Phosphatidylinositol-specific phospholipase C of Bacillus
Cereus.
Cloning, sequencing, and relationship to other phospholipases
J. Bacteriol. 171, 6077-6083 (1989)
90036694
COMMENT
Draft entry and computer-readable sequence for [1] kindly submitted
by K.Andreas, 03-OCT-1989.
FEATURES
Location/Qualifiers
1..2017
/organism="Bacillus cereus"
/db_xref="taxon:1396"
836..1825
/note="phospholipase C precursor"
/codon_start=1
/transl_table=11
/protein_id="AAA22665.1"
/db_xref="GI:143340"
/translation="MSKKLLIKLFCSTIFITFVFALHDKRVVAASSVNELENSKW
MQIPDNIPLARISIPGTHSDGTFKLQNP:KQVGMTOEYDFRYOMDHGAR:FD:RGR
LTDNDITVLHGGPLYLYVTLHEF:INEAKOFLKDNPSSETI:IMSLKKEYEDMKGAGSFS
STPEKNYFVDPIFLKTEGNIKLGDAKGI:VLLKRYSGSNEGGTNNIT:TWPNIT:FTTT
VYQNVAVTQDQKYKYNDEKYSIKDTMDTMMNSDNLNHY:INT:SI:SSGGT:AWNSP
YYIASYINPEIANV:IKQKNP:TRVGW:IQDY:INEKWSPLLYQEV:IRANKSI:IKP"
836..928
/note="phospholipase C signal peptide"
929..1822
/product="phospholipase C"
1839..1865
/note="pot. transcription termination signal: putative"
BASE COUNT 797 a 264 c 364 g 592 t
ORIGIN
1 ACAATTTTCCACGA 15
|||||
Db 1404 ACAATTTTCCACGA 1390

Query Match 100.0%; Score 15; DB 1; Length 2017;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAATTTTCCACGA 15
|||||
Db 1404 ACAATTTTCCACGA 1390
```

RESULT 4

BTPIPLC1/c
 LOCUS 2254 bp DNA linear BCT 22-MAR-1995
 DEFINITION Bacillus thuringiensis gene for phosphatidylinositol-specific phospholipase C PI-PLC (EC 3.1.4.10)
 X14178
 ACCSSION X14178.1 GI:40323
 VERSION 1
 KEYWORDS 1-phosphatidylinositol phosphodiesterase;
 phosphatidylinositol-specific phospholipase C.
 SOURCE Bacillus thuringiensis.
 ORGANISM Bacillus thuringiensis
 Bacteria; Firmicutes; Bacillus/Clostridium group;
 Bacillus/staphylococcus group; Bacillus; Bacillus cereus group.
 REFERENCE 1 (bases 1 to 2254)
 AUTHORS Lechner M.
 TITLE Direct Submission
 JOURNAL Submitted (24-JAN-1989) Lechner M., University Tuebingen, Mikrobielle Genetik, Auf der Morgenstelle 28, D-7400 Tuebingen, FRG
 REFERENCE 2 (bases 214 to 1592)
 AUTHORS Lechner M., Kupke T., Stefanovic S. and Gotz F.
 TITLE Molecular characterization and sequence of phosphatidylinositol-specific phospholipase C of Bacillus thuringiensis
 JOURNAL Mol. Microbiol. 3 (5), 621-626 (1989)
 MEDLINE 89343648
 COMMENT Data kindly reviewed (15-FEB-1990) by Lechner M.
 FEATURES
 Location/Qualifiers
 1..2254
 /organism="Bacillus thuringiensis"
 /strain="DSM 2046"
 /db_xref="taxon:1428"
 /clone_lib="pCA44."
 441..447
 /note="put. rRNA binding site"
 454..1443
 /note="PI-PLC precursor (AA -31 to 298)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA432378.1"
 /db_xref="GI:40324"
 /db_xref="SWISS-PROT:P08954"
 /translation="MSNKKLILKLFICSTIFITVFALHDKRVVAASSVNELENSKV
 MQIPIDNPLARIPIGTHDSGTFLKQNPDKVGMTOEDFYQMDHGRIPDINGR
 LTDDNTIVLHGPLYLVVTLRETFNEAKOFKADNPSEIIMSLKEDYDKAGSGFS
 STFEKNYFVDELTKETGNILKGDARKVILKRYGSGNESGGYNFYPDNETFTT
 VNQNVVQDKYKXVYDEKVKIKIDMETENSEDLNHLIYNFTSLSGGTAMNSP
 YYASYINPEANDIKQNPTRVGWVIQDYINEXWSPILYQEVIRANKSLIKE"
 sig_peptide 454..546
 mat_peptide 547..1440
 /product="mature PI-PLC (1-298)"
 BASE COUNT 844 a 271 c 430 g 709 t
 ORIGIN

Query Match 100.0%; Score 15; DB 1; Length 2254;
 Best Local Similarity 100.0%; Pred. NO. 4.5e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAATTTTCCACGA 15

|||||
 Db 1022 ACAATTTTCCACGA 1008

RESULT 5

AX346416/c
 LOCUS 15399 bp DNA linear PAT 01-FEB-2002
 DEFINITION Sequence 1487 from Patent WO0200928.
 ACCESSION AX346416
 VERSION AX346416.1 GI:18494302
 KEYWORDS synthetic construct.
 SOURCE synthetic construct.
 ORGANISM artificial construct.
 REFERENCE 1 (sites)

AUTHORS
 TITLE
 JOURNAL

FEATURES
 source

BASE COUNT 4553 a 156 c 3143 g 7547 t
 ORIGIN

Query Match 100.0%; Score 15; DB 6; Length 15399;
 Best Local Similarity 100.0%; Pred. NO. 3.5e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAATTTTCCACGA 15

|||||
 Db 13487 ACAATTTTCCACGA 13473

RESULT 6
 AC006717/c

LOCUS AC006717 27722 bp DNA linear INV to AKI 2001
 DEFINITION Caenorhabditis elegans cosmid Y14H12A, complete sequence.
 AC006717
 VERSION AC006717.1 GI:4263221
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 27722)

AUTHORS The C. elegans Sequencing Consortium.
 TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium
 JOURNAL Science, 282 (5396), 2012-2018 (1998)
 MEDLINE 99059613

REFERENCE 2 (bases 1 to 27722)

AUTHORS Rohlfing T. and Wohldmann P.
 TITLE The sequence of C. elegans cosmid Y14H12A
 JOURNAL Unpublished

REFERENCE 3 (bases 1 to 27722)

AUTHORS Waterston R.
 TITLE Direct Submission
 JOURNAL Unpublished

REFERENCE 4 (bases 1 to 27722)

AUTHORS Waterston R.H.
 TITLE Direct Submission
 JOURNAL Submitted (23-FEB-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 27722)

AUTHORS Waterston R.H.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 6 (bases 1 to 27722)

AUTHORS Waterston R.
 TITLE Direct Submission
 JOURNAL Submitted (24-MAR-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE 7 (bases 1 to 27722)

AUTHORS Waterston R.
 TITLE Direct Submission
 JOURNAL Submitted (09-AUG-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA

REFERENCE 8 (bases 1 to 27722)

AUTHORS Waterston R.
 TITLE Direct Submission
 JOURNAL Submitted (30-AUG-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE 9 (bases 1 to 27722)

AUTHORS Waterston R.
 TITLE Direct Submission
 JOURNAL Submitted (30-AUG-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE 10 (bases 1 to 27722)

AUTHORS Waterston R.
 TITLE Direct Submission
 JOURNAL Submitted (30-AUG-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE 11 (bases 1 to 27722)

AUTHORS Waterston R.
 TITLE Direct Submission
 JOURNAL Submitted (30-AUG-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA

COMMENT

Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: rwnematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).
Location/Qualifiers

FEATURES

source

```
1. 27722
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="II"
/clone="Y14H12A"
/size=5735
/gene="Y14H12A.1"
join(5232..5410,5457..5607,5655..5735)
/gene="Y14H12A.1"
/note="coded for by the following C. elegans cDNAs:
yk233d2.5, yk369b9.5, yk440c9.3, yk440c9.5"
/codon_start=1
/product="Hypothetical protein Y14H12A.1"
/protein_id="AAK84582.1"
/db_xref="GI:15145422"
/translation="MKKMDIMLVLCSTVIAHLIFLAFVLPYWNFTSPSSIMTIM
AFFLAIVSICVIFSLICIMHSSWSKIGVITFVWAFITGVLOASFFVWVVSOT
GSEFDGISTIGICAYFALVSRADOKPTSQLGHN"
```

BASE COUNT 8633 a 5108 c 4940 g 9041 t
ORIGIN
Query Match 100.0%; Score 15; DB 3; Length 27722;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ACAATTTTCCACGA 15
|||||
Db 18097 ACAATTTTCCACGA 18083

RESULT 7
AC009745 38199 bp DNA linear INV 21-APR-2001
LOCUS
DEFINITION
Drosophila melanogaster, chromosome 2L, region: 39D-39D, BAC clone
BACR01H19, complete sequence.
AC009745
ACCESSION
VERSION
KEYWORDS
SOURCE
HTG
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephyraoidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 38199)
AUTHORS
Ceiniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,

Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Anderson,R.C.,
Rogers,Y., An,H., Baldwin,D., Bantz,J., Beeson,K.Y., Busam,B.A.,
Carlson,J.M., Center,A., Chape,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Dou,L.E., Doyle,C., Dresnick,D., Farfan,B.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,K.A.,
Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshier,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nurnberg,J.,
Pacile,J., Paragov,V., Park,S., Patel,S., Pfeiffer,B., Scheerlin,P.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheerlin,P.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 2L, region 39D-39D;
Unpublished

TITLE
JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 38199)
Ceiniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,K., Blazek,P.G.,
Rutenhoff,C., Champe,M., Chavez,C., Chew,M., Chesliska,I.,
Doyle,C.M., Farfan,B.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshier,A.R., Moshier,M., Nixon,K., Pacile,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Svirskas,R.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zierahn,L.L. and
Rubin,G.M.

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (31-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Apr 21, 2001 this sequence version replaced gi:6468061.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu.

FEATURES

source

```
1. 38199
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="39D-39D"
/clone="BACR01H19 (D1039)"
/clone_lib="RPC1-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial BAC) in
BACE3.6)"
BASE COUNT 11950 a 6893 c 7231 g 12125 t  
ORIGIN
```

Query Match 100.0%; Score 15; DB 3; Length 38199;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACAATTTTCCACGA 15
|||||

Db 23469 ACAATTTTCCACGA 23483

RESULT 8

AC018103

LOCUS

DEFINITION

AC018103 Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC018103.1 GI:6553088
HTG; HTGS-PHASE2.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

REFERENCE
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 40018)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10213388 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

CDS
 join(7576..8273,8635..9487,9997..10138,10305..10429,10596..10824)
 /gene="K07D4.7"
 /note="contains similarity to pfam families PF00621 (RhoGEF domain, score=65.2, E=1.4e-15, N=1), PF00018 (SH3 domain, score=44.8, E=1.9e-09, N=1) and PF00169 (PH domain, score=21.6, E=0.00012, N=1); coded for by the following C. elegans cDNAs: yk356c12.3, yk356c12.5"
 /codon_start=1
 /product="Hypothetical protein K07D4.7b"
 /protein_id="AAC26284.1"
 /db_xref="GI:3319429"
 /translation="MSDTEKIVAPVAVGHPLPDDSWTPTSAEAIODEPSPDEEYKRO YTCWPNPGLPIQHYMLEAVAVADDEVPLEEVKEEFPVLPAAVHEVVALKPAIR RESSIDSGRGADCSITTSALTSNTSMRRLVGTNFGSORSLSWCELPVRAAGL LKIDECERQAEVFTSEASYLSNLVLTHTFMAPSPOMLASKALSLSVSDSRK HLFNSNFAVDCSERLLCDLEENLILDDICDILSDHFKHFEVYKYSNQVYQ DTRRLKLTENPGFLSAVORLEENKOCQGLMRSFLMLPMQVRYPLLYLAILDRIT TTDRVKTATDALSSNRVROCNAGRWERTEQLLDIDRELYIKDADLKRIPLVSN SRLVKKGVLTQLVERRSSNLIQSRQARTLHVFLFSDMIMTKKLNQGVFKQYAA RFVDMQTEPNPKIPIGATSNLVRPHFLCTLMRNARDKCTELLSSADSETDRE RWSAVRPPTSTNPEKIYAEWDCQALVHVAYVPOADELOLIIGDSLILRKPQGW LYGERAGDGLGSPSSVQVAILNDHTRANNVQRRLIIQASNGHGNASPSSTRSGG MPLMDRLRMSNPKTYFQSGTGL"
 /complement(11882..13756)
 /gene="K07D4.6"
 complement(join(11882..12345,12395..12471,12531..12655,13145..13291,13343..13502,13560..13756))
 /gene="K07D4.6"
 /codon_start=1
 /product="Hypothetical protein K07D4.6"
 /protein_id="AAC26290.1"
 /db_xref="GI:3319435"
 /translation="MSFTSHFLFLAYLTSSGSSDGRAPPCOLPPTNLLSINS RELAKIWAGFHSSTSAKCELOACTKELLAGLAEDMKMKILLDYDRPVSVEERR TNFLAGSEQTKMGSGINSRSIHTDEYDLNNAWEHLDDAMNVEQVETFSQQR KNKSKLAVDNLSSEARLAHRESLRLLKQDVYSNSARKEAGLRAKCPRGSI YASILDABEQALQSDVDRIPLFTTHSSSTAVPGTETSETPETTTSTTTT EPIITSTIQTITVTSTISTSTSTTTTPTTTTSTTTSDLLAWKITVTPP PPTTEGLSEFEIEKSNNTVPLTPSRPHKVASMEFVADPEKLTLLKQIFGHKND"
 /complement(15537..18903)
 /gene="K07D4.5"
 complement(join(15537..15684,16550..16649,17449..17625,18608..18734,18829..18903))
 /gene="K07D4.5"
 /codon_start=1
 /product="Hypothetical protein K07D4.5"
 /protein_id="AAC26289.1"
 /db_xref="GI:3319434"
 /translation="MLHPTSFVNSYKKNYLEWRNPITTHNAEAVNRICQSIK TDFKTLFLGDUETSEFYDIDRAQKFPKKNLKNKKNFLKLEKIEIDILAENI ISKLRLKTSFKNKFNFEIVLNFDAHFWDGDAYGLFGFVSPDFVSGAPNRNQIW FLRNDGLYLSEHFNRAREWQWMPPLDSYATESIDQVDNSPLPAWVPY"
 /complement(26436..27703)
 /gene="K07D4.4"
 complement(join(26436..26637,27070..27252,27322..27496,27664..27703))
 /gene="K07D4.4"
 /codon_start=1
 /product="Hypothetical protein K07D4.4"
 /protein_id="AAC26288.1"
 /db_xref="GI:3319433"
 /translation="MPIKRNIVVCKREYFONILLHFNQIILGHKMQSFCPIILFF TFVLCVKSODEVFIIMDHTSPDEQGGYTRPKFRKFLIIFENLEIQVRNLKIVA LKILESEKPEFLWLVFYDYHTSENFALFSHIVALRPDYTWKWNVLKTVFDS RQTEKRGPRTRIEEVPAKLPNLSRLLNFGYRI"
 /complement(34018..38108)
 /gene="K07D4.8"
 join(34018..34074,35411..35597,35879..35978,36270..36392,36438..36531,36581..36661,36750..36865,37907..38108)
 /gene="K07D4.8"
 /note="contains similarity to human gamma-interferon inducible lysosomal thiol reductase (SUB 4-104)"

Query Match 100.0%; Score 15; DB 3; Length 44334;
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ACAATTTTCCACCA 15
 |||||
 DB 1225 ACAATTTTCCACCA 1211
 |||||
 RESULT 10
 AB026651/c
 LOCUS
 DEFINITION
 Arabidopsis thaliana genomic DNA, chromosome 5, Pl clone:MQD19,
 ACCESSION
 AB026651 BAO00015
 VERSION
 AB026651.1 GI:4757407
 KEYWORDS
 SOURCE
 Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P-1
 clone:MQD19.
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 1 (sites)
 Kaneko,T., Katoh,T., Asamizu,E., Sato,S., Nakamura,Y., Kotani,H.
 and Tabata S.
 Structural analysis of Arabidopsis thaliana chromosome 5. XI
 Unpublished
 2 (bases 1 to 87286)
 Nakamura,Y.
 Direct Submission
 Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
 Institute, Department of Plant Gene Research; 1532-3, Yana,
 Kisarazu, Chiba 252-0812, Japan (E-mail:ynakam@kazusa.or.jp,
 Tel:81-438-52-3935, Fax:81-438-52-3934)
 Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see
 http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MQD19
 Genes with similarity to proteins in the databases are described in
 'product' or 'note' qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Graal
 (Informatics Group, Oak Ridge National Laboratory;
 http://complib.ornl.gov/Graal-1.3/),
 GENSCAN (Chris Burge, MIT, http://ccr.mit.edu/GENSCAN.html),
 NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
 Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
 SplicePredictor (Volker Brendel, Stanford University,
 http://genome1.ool.iastate.edu/cgi-bin/sp.cgi).
 Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
 http://genome.wustl.edu/eddy/tRNAscan-SE/).
 This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.
 The 5' clone is MQD24 and the 3' clone is F686.
 FEATURES
 Location/Qualifiers
 1..87286
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="MQD19"


```

CDS
    join(40794..41579,42574..43377)
    /note="gene_id:MQD19.11"
    /codon_start=1
    /evidence=not_experimental

Query Match      100.0%; Score 15; DB 8; Length 87286;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAATTTCACGCA 15
    |||
Db 71761 ACAATTTCACGCA 71747

RESULT 11
LOCUS      MUSTCRA          94547 bp    DNA    linear    ROD 27-APR-1993
DEFINITION Mouse T-cell receptor alpha/delta chain locus.
ACCESSION M64239
VERSION   M64239.1 Gi:201744
FEATURES             1..94547
     source            Mus musculus
     ORGANISM          Mus musculus
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Rodentia; Sciurgnathi; Muridae; Murinae; Mus.
     Chien, Y.-H., Becker, D.M., Lindsten, T., Okamura, M., Cohen, D.I. and
     Davis, M.M.
     A third type of murine T-cell receptor gene
     Nature 312, 31-35 (1984)
     85036634
     2 (sites)
     Saito, H., Kranz, D.M., Takagaki, Y., Hayday, A.C., Eisen, H.N. and
     Tonegawa, S.
     A third rearranged and expressed gene in a clone of cytotoxic T
     lymphocytes
     Nature 312, 36-39 (1984)
     85036635
     3 (sites)
     Patten, P., Yokota, T., Rothbard, J., Chien, Y.-H., Arai, K.-I. and
     Davis, M.M.
     Structure, expression and divergence of T-cell receptor beta-chain
     variable regions
     Nature 312, 40-46 (1984)
     85036636
     4 (sites)
     Arden, B., Klotz, J.L., Siu, G. and Hood, L.E.
     Diversity and structure of genes of the alpha family of mouse
     T-cell antigen receptor
     Nature 316, 783-787 (1985)
     85296324
     5 (sites)
     Hayday, A.C., Diamond, D.J., Tanigawa, G., Heilq, J.S., Folsom, V.,
     Saito, H. and Tonegawa, S.
     Unusual organization and diversity of T-cell receptor alpha-chain
     genes
     Nature 316, 828-832 (1985)
     85296331
     6 (sites)
     Becker, D.M., Patten, P., Chien, Y.-H., Yokota, T., Eshhar, Z.,
     Giedlin, M., Gascoigne, N.R.J., Goodnow, C., Wolf, R., Arai, K. and
     Davis, M.M.
     Variability and repertoire size of T-cell receptor V-alpha gene
     segments
     Nature 317, 430-434 (1985)
     86014379
     7 (sites)
     Baer, R.J., Lefranc, M.-P., Minowada, J., Forster, A., Stinson, M.A. and
     Rabbitts, T.H.
     Organization of the T-cell receptor alpha-chain gene and
     rearrangement in human T-cell leukaemias
     Mol. Biol. Med. 3, 265-277 (1986)

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
Nucleotide sequence analysis of 95 kb near the 3' end of the murine
T-cell receptor alpha/delta chain locus: Strategy and methodology
Genomics 13, 1198-1208 (1992)
Authorin kindly submitted by R.K. Wilson 1-26-91.
location/Qualifiers
1..94647
/organism="Mus musculus"
/strain="BA1B/C"
/sub_species="domesticus"
/db_xref="taxon:10090"
/tissue_type="liver"
2454..2452
/note="putative recombination recognition sequence;
putative"
2475..2481
/note="putative recombination recognition sequence";
putative"
2482..2540
/partial

misc_signal
misc_signal
CDS

```

```

86284199
8 (sites)
Dembic, Z., Haas, W., Weiss, S., McCubrey, J., Kiefer, H., Von
Roehmer, H. and Steinmetz, M.
Transfer of specificity by murine alpha and beta T-cell receptor
genes
Nature 320, 232-238 (1986)
86175007
9 (sites)
Winoto, A., Urban, J., Lan, N., Gorman, J., Hood, L.E. and Hansburg, D.
Predominant use of a V-alpha gene segment in mouse T-cell receptors
for cytochrome C
Nature 324, 679-682 (1986)
87090381
10 (sites)
Hochgeschwender, U., Simon, H.-G., Weltzien, H.U., Bartels, F.,
Becker, A. and Epplein, J.T.
Dominance of one T-cell receptor in the H-2Kb/TNP response
Nature 326, 307-309 (1987)
87144563
11 (sites)
Chien, Y.-H., Iwashima, M., Kaplan, K.B., Elliott, J.F. and Davis, M.M.
A new T-cell receptor gene located within the alpha locus and
expressed early in T-cell differentiation
Nature 327, 677-682 (1987)
87258184
12 (sites)
Iwashima, M., Green, A., Davis, M.M. and Chien, Y.-H.
Variable region (V-delta) gene segment most frequently utilized in
adult thymocytes is 3' of the constant (C-delta) region
Proc. Natl. Acad. Sci. U.S.A. 85, 8161-8165 (1988)
89042145
13 (sites)
Winoto, A. and Baltimore, D.
Alpha-beta lineage-specific expression of the alpha T cell receptor
gene by nearby silencers
Cell 59, 649-655 (1989)
90058647
14 (sites)
Winoto, A. and Baltimore, D.
A novel, inducible and T cell-specific enhancer located at the 3'
end of the T cell receptor alpha locus
EMBO J. 8, 729-733 (1989)
89251603
15 (sites)
Takehita, S., Toda, M. and Yamagishi, H.
Excision products of the T cell receptor gene support a progressive
rearrangement model of the Alpha/Delta locus
EMBO J. 8, 3261-3270 (1989)
90059917
16 (bases 1 to 94647)
Wilson, R.K., Koop, B.F., Chen, C., Halloran, N., Sciammis, K. and
Hood, L.
Nucleotide sequence analysis of 95 kb near the 3' end of the murine
T-cell receptor alpha/delta chain locus: Strategy and methodology
Genomics 13, 1198-1208 (1992)
92372016
Authorin kindly submitted by R.K. Wilson 1-26-91.
location/Qualifiers
1..94647
/organism="Mus musculus"
/strain="BA1B/C"
/sub_species="domesticus"
/db_xref="taxon:10090"
/tissue_type="liver"
2454..2452
/note="putative recombination recognition sequence;
putative"
2475..2481
/note="putative recombination recognition sequence";
putative"
2482..2540
/partial

misc_signal
misc_signal
CDS

```

```

/note="TCR delta chain joining region gene segment 2';
putative"
/codon_start=1
/label=J-delta_2
/protein_id="AA040346.1"
/db_xref="GI:201745"
/translation="LLGHFTDVFVNWHRALCGAP"
2541..8568
/note="TCR delta chain J2-C intervening DNA"; putative"
8569..8847
/note="TCR delta chain constant region gene, exon 1';
putative"
/citation=[11]
/label=C-delta_exon1
8848..9508
/note="putative"
9509..9574
/note="TCR delta chain constant region gene, exon 2';
putative"
/citation=[11]
/label=C-delta_exon2
9575..9934
/note="putative"
9935..10051
/note="TCR delta chain constant region gene, exon 3';
putative"
/citation=[11]
/label=C-delta_exon3
10052..11191
/note="putative"
11192..11817
/note="3' untranslated region"; 'TCR delta chain constant
region gene, exon 4'; putative"
/citation=[11]
/label=C-delta_exon4
11812..11817
/note="putative"
14284..14292
/note="putative recombination recognition sequence";
putative"
14316..14322
/note="putative recombination recognition sequence";
putative"
14323..14669
/note="TCR delta chain variable region gene
segment(V-delta 5)'; 'reverse translational orientation';
putative"
/citation=[12]
/label=V-delta5_exon2
14670..14825
/note="putative"
14826..14862
/note="TCR delta chain variable region gene
segment(V-delta 5)'; 'reverse translational orientation';
putative"
/citation=[12]
/label=V-delta5_exon1
20624..20632
/note="putative recombination recognition sequence";
putative"
20645..20651
/note="putative recombination recognition sequence";
putative"
20652..20723
/partial
/note="TCR alpha chain joining region gene segment,
pseudogene (pseudol)"; 'pseudogene; contains 2
termination codons'; putative"
/citation=[15]
/codon_start=1

```

Query Match 100.0%; Score 15; DB 10; Length 94647;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;

```

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACAATTTCCACGA 15
|||||
DB 20855 ACAATTTCCACGA 20869

RESULT 12
AE008686_3
WPCOMMENT
Sequence split into 5 fragments LOCUS AE008686 Accession AE008686
Fragment Name Begin End
AE008686_0 1 110000
AE008686_1 100001 210000
AE008686_2 200001 310000
AE008686_3 300001 410000
AE008686_4 400001 461875
Continuation (4 of 5) of AE008686 from base 300001 (AE008686 Mus musculus 1 cell rec
Query Match 100.0%; Score 15; DB 10; Length 110000;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACAATTTCCACGA 15
|||||
DB 88083 ACAATTTCCACGA 88097

RESULT 13
AP003776
LOCUS
DEFINITION
Homo sapiens chromosome 11 clone RP11-209L12 map 11q, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
AP003776
ACCESSION
AP003776.1 GI:14530688
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens DNA, clone:RP11-209L12.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 152948)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (21-JUN-2001) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
----- Genome Center -----
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information -----
Center project name: Humbrafill
Center clone name: RP11-209L12
----- Summary Statistics -----
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 145988 bases at least Q40
Consensus quality: 151056 bases at least Q30
Consensus quality: 151725 bases at least Q20
Insert size: 151948; sum-of-contigs
Quality coverage: 5.98x in Q20 bases; sum-of-contigs

```

NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and the order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be.

preserved

1 35942 contig of 35942 bp in length
 36043 64596 contig of 28554 bp in length
 64597 88056 contig of 23360 bp in length
 88157 108127 contig of 19971 bp in length
 108228 121652 contig of 13425 bp in length
 121753 132902 contig of 11150 bp in length
 133003 139176 contig of 6174 bp in length
 139277 144517 contig of 5241 bp in length
 144618 147983 contig of 3366 bp in length
 148084 152287 contig of 4204 bp in length
 152388 152948 contig of 561 bp in length.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 11 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 35942: contig of 35942 bp in length
 35943 36042: gap of 100 bp
 36043 64596: contig of 28554 bp in length
 64597 64696: gap of 100 bp
 64697 88056: contig of 23360 bp in length
 88057 88156: gap of 100 bp
 88157 108127: contig of 19971 bp in length
 108128 108227: gap of 100 bp
 108228 121652: contig of 13425 bp in length
 121653 121752: gap of 100 bp
 121753 132902: contig of 11150 bp in length
 132903 133002: gap of 100 bp
 133003 139176: contig of 6174 bp in length
 139177 139276: gap of 100 bp
 139277 144517: contig of 5241 bp in length
 144518 144617: gap of 100 bp
 144618 147983: contig of 3366 bp in length
 147984 148083: gap of 100 bp
 148084 152287: contig of 4204 bp in length
 152288 152387: gap of 100 bp
 152388 152948: contig of 561 bp in length.

FEATURES

SOURCE

1..152948
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11q"
 /clone="RP11-209L12"

misc_feature

1..35942
 /note="assembly_fragment"
 36043..64596
 /note="assembly_fragment"
 64697..88056
 /note="assembly_fragment"
 88157..108127
 /note="assembly_fragment"
 108228..121652
 /note="assembly_fragment"
 121753..132902
 /note="assembly_fragment"
 133003..139176
 /note="assembly_fragment clone_end:T7 vector_side:left"
 139277..144517
 /note="assembly_fragment"
 144618..147983
 /note="assembly_fragment"
 148084..152287
 /note="assembly_fragment"
 152388..152948
 /note="assembly_fragment clone_end:SP6 vector_side:left"

BASE COUNT 42013 a 33863 c 34632 g 41440 t 1000 others
 ORIGIN

Query Match

Best Local Similarity 100.0%; Score 15; DB 2; Length 152948;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACAATTTCCACGCA 15

Db 139531 ACAATTTCCACGCA 1,545

RESULT 14

AC022208/c

LOCUS

DEFINITION

AC022208

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

On Aug 17, 2000 this sequence version replaced at:7523889.

Center: Washington University Genome Sequencing Center

Center code: WUCSC

Web site: http://genome.wustl.edu/gsc/index.shtml

Center project name: H_NH055G19

Sequencing vector: M13; 100%

Chemistry: Dye-terminator Big Dye; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 163053 bases at least 040

Consensus quality: 164948 bases at least 030

Consensus quality: 165791 bases at least 020

Insert size: 174000; agarose-fp

In. rt size: 169159; sum-of-contigs

Quality coverage: 4.85 in Q20 bases; agarose-fp

Quality coverage: 5.04 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 14 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 2787: contig of 2787 bp in length

* 2788 2887: gap of unknown length

* 2888 5270: contig of 2383 bp in length

* 5271 5370: gap of unknown length

* 5371 10592: contig of 5222 bp in length

* 10593 10692: gap of unknown length

* 10693 16160: contig of 5468 bp in length

* 16161 16260: gap of unknown length

* 16261 22457: contig of 6197 bp in length

* 22458 22557: gap of unknown length

* 22558 29902: contig of 7345 bp in length

* 29903 30002: gap of unknown length

* 30003 39328: contig of 9326 bp in length.
 * 39329 39428: gap of unknown length
 * 39429 51983: contig of 12555 bp in length
 * 51984 52083: gap of unknown length
 * 52084 66277: contig of 14194 bp in length
 * 66278 66377: gap of unknown length
 * 66378 81611: contig of 15234 bp in length
 * 81612 81711: gap of unknown length
 * 81712 100091: contig of 18380 bp in length
 * 100092 100191: gap of unknown length
 * 100192 118824: contig of 18633 bp in length
 * 118825 118925: gap of unknown length
 * 118926 141074: contig of 22150 bp in length
 * 141075 141175: gap of unknown length
 * 141176 169400: contig of 28226 bp in length.

FEATURES
 Location/Qualifiers
 1..169400
 Source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /clone="RP11-555G19"

misc_feature
 1..2787
 /note="assembly_name:Contig4
 clone_end:17
 vector_side:right"

misc_feature
 2888..5270
 /note="assembly_name:Contig5"

misc_feature
 5371..10592
 /note="assembly_name:Contig6"

misc_feature
 10693..16160
 /note="assembly_name:Contig7"

misc_feature
 16261..22457
 /note="assembly_name:Contig8"

misc_feature
 22558..29902
 /note="assembly_name:Contig9"

misc_feature
 30003..39328
 /note="assembly_name:Contig10"

misc_feature
 39429..51983
 /note="assembly_name:Contig11"

misc_feature
 52084..66277
 /note="assembly_name:Contig12"

misc_feature
 66378..81611
 /note="assembly_name:Contig13"

misc_feature
 81712..100091
 /note="assembly_name:Contig14
 clone_end:SP6"

misc_feature
 vector_side:left
 100192..118824
 /note="assembly_name:Contig15"

misc_feature
 118925..141074
 /note="assembly_name:Contig16"

misc_feature
 141175..169400
 /note="assembly_name:Contig17"

BASE COUNT 47133 a 37228 c 38625 g 45097 t 1317 others
 ORIGIN

Query Match 100.0%; Score 15; DB 2; Length 169400;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAATTTCCACGA 15
 |||
 Db 149720 ACAATTTCCACGA 149706

RESULT 15
 AP003062
 LOCUS
 DEFINITION Homo sapiens genomic DNA, chromosome 11q clone:RP11-555G19,
 complete sequences.
 ACCESSION AP003062
 VERSION AP003062.2 GI:18307729
 KEYWORDS HTG.

SOURCE Homo sapiens DNA, clone:RP11-555G19.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 169556)
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seon,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (25-DEC-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suenro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 COMMENT On Jan 23, 2002 this sequence version replaced qi:11995050.

FEATURES
 Location/Qualifiers
 1..169556
 Source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11q"
 /clone="RP11-555G19"

BASE COUNT 46187 a 37659 c 38846 g 46864 t
 ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 169556;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAATTTCCACGA 15
 |||
 Db 108399 ACAATTTCCACGA 108413

Search completed: October 17, 2002, 10:51:59
 Job time : 1204 secs

GenCore version 5.1.3

Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 09:32:18 ; Search time 1196 Seconds

(without alignments)
169.276 Million cell updates/sec

Title: US-09-820-203A-1

Perfect score: 15
Sequence: 1 agtatgggaatgag 15Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	100.0	177	9	AW294916 UI-H-BI2-
2	15	100.0	187	9	AA885068 am11006.s
3	15	100.0	210	9	BE047631 BB047631
4	15	100.0	270	10	R08860 yf21f08.s1
5	15	100.0	320	10	R77770 vi77e07.s1
6	15	100.0	325	9	AA972014 op84b05.s
7	15	100.0	345	10	T77464 yd3d04.s1
8	15	100.0	364	9	AW242622 xm01e07.x
9	15	100.0	384	9	AU185018 AU185018
10	15	100.0	410	9	AI698694 wa80g01.x
11	15	100.0	410	12	AQ315159 RPI11-10
12	15	100.0	413	9	AI393449 tg45d08.x
13	15	100.0	435	9	AI127080 qb97f10.x
14	15	100.0	440	9	AI654689 wb49a08.x
15	15	100.0	448	9	AW027474 wt95c08.x
16	15	100.0	462	10	BI878210 fl75h12.x
17	15	100.0	532	12	AZ010661 RPI-23-3

c	18	15	100.0	684	12	BH353141
c	19	15	100.0	697	12	AZ117932
c	20	15	100.0	825	12	BH151494
c	21	15	100.0	891	10	BI462410
c	22	14.6	97.3	950	12	CNS040DC
c	23	14	93.3	103	12	AZ768976
c	24	14	93.3	200	10	BF169826
c	25	14	93.3	248	10	EG586404
c	26	14	93.3	279	9	BE254943
c	27	14	93.3	329	9	BB115880
c	28	14	93.3	335	9	BB099407
c	29	14	93.3	344	10	H07923
c	30	14	93.3	361	9	AA935079
c	31	14	93.3	363	9	AW736989
c	32	14	93.3	416	10	EG040654
c	33	14	93.3	424	9	BB750771
c	34	14	93.3	435	12	BH187359
c	35	14	93.3	435	12	CNS07RNN
c	36	14	93.3	436	10	BJ046122
c	37	14	93.3	453	9	AW114313
c	38	14	93.3	457	12	AQ672800
c	39	14	93.3	461	10	R98219
c	40	14	93.3	471	9	BB85029
c	41	14	93.3	476	12	AZ229693
c	42	14	93.3	480	12	BH350887
c	43	14	93.3	483	10	BF410024
c	44	14	93.3	493	9	BB854947
c	45	14	93.3	496	9	AW280565

ALIGNMENTS

AW294916 177 bp mRNA linear EST 16-JAN-2000
 UI-H-BI2-ahr-e-03-0-UI.s1 NCI_CGAP_Sub4 Homo sapiens cDNA clone
 IMAGE:2727653 3', mRNA sequence.

ACCESSION AW294916
 VERSION AW294916.1
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 177)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

Contact: Robert Straube, Ph.D.

Email: cgapbs-r@mail.nih.gov

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:

NCI-CGAP clone distribution information can be found through the

I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward

POLYA=Yes.

Location/Qualifiers

1..177

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2727653"

/clone_lib="NCI_CGAP_Sub4"

/lab_host="DHL0B (Life Technologies)"

/note="Vector: p773D-pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; The

NCI-CGAP Sub4 library is a subtracted library derived from

the NCI-CGAP Sub2 library which is a subtracted library

derived from the NCI-CGAP Subl library, which is a

subtracted library derived from B1. B1 constitutes a mixture of 21 normalized or subtracted NCI_CGAP libraries: NCI_CGAP_Co4, NCI_CGAP_Pr22, NCI_CGAP_Pt28, NCI_CGAP_Co10, NCI_CGAP_Co16, NCI_CGAP_Kid5, NCI_CGAP_Kid12, NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2, NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_CLL1, NCI_CGAP_Le12, NCI_CGAP_Brn23, NCI_CGAP_Lu5, NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6, NCI_CGAP_Brn25. These 21 libraries were pooled and a single-stranded DNA preparation of the resulting mixture was used as a tracer in a subtractive hybridization with a driver whose composition is detailed below:

NCI_CGAP_Kid3 pool 1 : LLAM 3334-3337, 3682-3683, 3798-3803 (IMAGE CloneIDs 1322376-1323911, 1456008-1456775, 1500552-1502855) NCI_CGAP_Kid5 pool 1 : LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE CloneIDs 1323912-1325831, 1471368-1472903, 1492104-1493255) NCI_CGAP_Lu5 pool 1 : LLAM 3575-3582, 3851-3854 (IMAGE CloneIDs 1414920-1417991, 1520904-1522439) NCI_CGAP_GC4 pool 1 : LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE CloneIDs 1257096-1258631, 1469064-1470983, 1475592-1476743) NCI_CGAP_Pr22 pool 1 : LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE CloneIDs 985608-986759, 1101182-1101959, 1217928-1220615) NCI_CGAP_Co10 pool 1 : LLAM 2844-2853, 2871-2872 (IMAGE CloneIDs 1057416-1061255, 1144584-1145351)

) Subtraction was performed as previously described [Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery. Genome Research 6, 791-806.]

TAG_LIB=NCI_CGAP_GC4
TAG_TISSUE=germ Cell
TAG_SEQ=AAATC"

source

1. 320
 /organism="Homo sapiens"
 /db_xref="CDB:563067"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:145284"
 /sex="Female"
 /dev_stage="Placenta obtained at birth (full term)"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAATTCGCCGCGAGCAATTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."
 120 a 38 c 74 g 87 t 1 others

BASE COUNT

Query Match 100.0%; Score 15; DB 10; Length 320;
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACTATGGGGAATGAG 15

Db 114 AGTATGGGGAATGAG 128

RESULT 6

AA972014

LOCUS

DEFINITION op84h05.sl Soares_NFL_T_GBC-sl Homo sapiens cDNA clone IMAGE:1583577 3', mRNA sequence. EST 07-JUL-1998

ACCESSION

AA972014

VERSION

AA972014.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 325)

AUTHORS

NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 This clone is available royalty-free through LINDL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 618 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 318.
 Location/Qualifiers
 1. 325

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1583577"
 /clone_lib="Soares_NFL_T_GBC-sl"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHR, and B-cell NCI-CCAP-GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1 M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

ORIGIN 121 a 41 c 75 g 88 t

Query Match 100.0%; Score 15; DB 9; Length 325;
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGTATGGGGAATGAG 15

Db 115 AGTATGGGGAATGAG 129

RESULT 7

T77464

LOCUS

DEFINITION yd73d04.sl Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:113863 3', mRNA sequence. EST.

ACCESSION

T77464

VERSION

T77464.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 345)

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.

TITLE

The WashU-Merck EST Project

JOURNAL

Unpublished (1995)

COMMENT

Other ESTs: yd73d04.r1
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 1054
 High quality sequence stops: 332
 This clone is available royalty-free through LINDL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1054 Std Error: 0.00
 Seq primer: -21ml3
 High quality sequence stop: 332.
 Location/Qualifiers
 1. 345

FEATURES

source

/organism="Homo sapiens"
 /db_xref="CDB:469480"
 /db_xref="taxon:9606"
 /clone="IMAGE:113863"
 /clone_lib="Soares fetal liver spleen INFLS"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

ORIGIN 123 a 46 c 79 g 97 t

Query Match 100.0%; Score 15; DB 10; Length 345;
 Best Local Similarity 100.0%; Pred. No. 1.7e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1

ACTATGGGGAATGAG 15

|||||

Db 115 AGTATGGGAATGAG 129

RESULT 8

AW242622

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Query Match

Best Local Similarity

Matches 15; Conservative

0; Mismatches

0; Indels

0; Caps

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

AW242622 364 bp mRNA linear EST 14-DEC-1999

NCI-CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2692452 3',

mRNA sequence.

AW242622

AW242622.1 GI:5576376

human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 364)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapsb@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Hammer-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -400P from Gibco

High quality sequence stop: 348.

Location/Qualifiers

1..364

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2692452"

/clone_lib="NCI-CGAP_Kid11"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Plasmid DNA from the normalized library NCI-CGAP_Kid3 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(cloneIDs 1322376-1323911, 1456007-1456775, and

1500552-1502855). Subtraction by Bento Soares and M.

Fatima Bonaldo."

131 a 51 c 79 g 103 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 15; Conservative

0; Mismatches

0; Indels

0; Caps

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 384)

Sugita, Y., Oshida, T. and Oya, Y.

Human cDNA sequencing

Unpublished (2001)

Contact: Yuji Sugita

Genex Research, Inc.

907 Nogawa, Miyamae-ku, Kawasaki, Kanagawa 216-0001, Japan

Tel: 81-44-797-2281

Fax: 81-44-797-2622

Email: syuji@genex.co.jp, URL: http://www.genex.co.jp.

Location/Qualifiers

1..384

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="B00768-020"

/clone_lib="Homo sapiens T-cell library (Sugita Y)"

/cell_type="T-cell"

128 a 57 c 87 g 112 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 15; Conservative

0; Mismatches

0; Indels

0; Caps

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 384)

Sugita, Y., Oshida, T. and Oya, Y.

Human cDNA sequencing

Unpublished (2001)

Contact: Yuji Sugita

Genex Research, Inc.

907 Nogawa, Miyamae-ku, Kawasaki, Kanagawa 216-0001, Japan

Tel: 81-44-797-2281

Fax: 81-44-797-2622

Email: syuji@genex.co.jp, URL: http://www.genex.co.jp.

Location/Qualifiers

1..384

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="B00768-020"

/clone_lib="Homo sapiens T-cell library (Sugita Y)"

/cell_type="T-cell"

128 a 57 c 87 g 112 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 15; Conservative

0; Mismatches

0; Indels

0; Caps

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

REFERENCE


```

source
1. .435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1708075"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGCATCTTTTCTTTTCTTTTCTTTT
3']],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M. Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHH19W."

BASE COUNT      144 a   63 c   99 g   129 t
ORIGIN

Query Match      100.0%; Score 15; DB 9; Length 435;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
|||||
Db 113 AGTATGGGAATGAG 127

RESULT 14
A1654689
LOCUS      w049a08.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2308982 3',
DEFINITION
ACCESSION  A1654689
VERSION    A1654689.1 GI:4738668
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 440)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 636 std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 414.
Location/Qualifiers
1. .440
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2308982"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP_GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."

FEATURES
source
1. .448
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2515214"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP_GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT      147 a   69 c   102 g   130 t
ORIGIN

Query Match      100.0%; Score 15; DB 9; Length 448;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
|||||
Db 116 AGTATGGGAATGAG 130

RESULT 15
AW027474
LOCUS      wt95c08.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2515214 3',
DEFINITION
ACCESSION  AW027474
VERSION    AW027474.1 GI:5886230
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 448)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. .448
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2515214"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP_GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT      147 a   69 c   102 g   130 t
ORIGIN

Query Match      100.0%; Score 15; DB 9; Length 448;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
|||||
Db 116 AGTATGGGAATGAG 130

```

```

source
1. .435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1708075"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGCATCTTTTCTTTTCTTTTCTTTT
3']],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M. Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHH19W."

BASE COUNT      144 a   63 c   99 g   129 t
ORIGIN

Query Match      100.0%; Score 15; DB 9; Length 435;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTATGGGAATGAG 15
|||||
Db 113 AGTATGGGAATGAG 127

RESULT 14
A1654689
LOCUS      w049a08.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2308982 3',
DEFINITION
ACCESSION  A1654689
VERSION    A1654689.1 GI:4738668
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 440)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 636 std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 414.
Location/Qualifiers
1. .440
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2308982"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA

```

Db 115 ACTATGGGCAATGAG 129

Search completed: October 17, 2002, 11:23:05
Job time : 1205 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 08:45:30 : Search time 145 Seconds
(without alignments)
177.612 Million cell updates/sec

Title: US-09-820-203A-2

Perfect score: 15

Sequence: 1 acaatttccacga 15

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	15	100.0	15399	24	Human immune syste
C 2	14	93.3	752	22	Human breast cance
C 3	14	93.3	7588	22	Human reproductive
C 4	14	93.3	7588	22	Human breast or ov
C 5	14	93.3	34185	21	Nucleotide sequenc
C 6	14	93.3	34185	21	Human breast cance
C 7	13.4	89.3	282	22	Novel human polynu
C 8	13.4	89.3	380	22	Human breast cell
C 9	13.4	89.3	450	22	Human breast cell

C 10	13.4	89.3	450	22	ABA56580	Human foetal liver
C 11	13.4	89.3	450	22	ABA26195	Probe #4661 for ge
C 12	13.4	89.3	450	22	AAK04728	Human brain expres
C 13	13.4	89.3	450	22	AAK30247	Human bone marrow
C 14	13.4	89.3	450	22	AAI14848	Probe #4781 for ge
C 15	13.4	89.3	450	22	AAI36207	Probe #4893 used t
C 16	13.4	89.3	450	22	AAI04634	Probe #4625 used t
C 17	13.4	89.3	497	21	AAK37564	Arabidopsis thalia
C 18	13.4	89.3	581	22	AAI25147	Human breast cance
C 19	13.4	89.3	609	22	AAK23908	Human ovarian PCR-
C 20	13.4	89.3	609	22	AAH82465	Human ovarian immo
C 21	13.4	89.3	622	22	AAH87886	Peppermint plant o
C 22	13.4	89.3	677	21	AAK59225	Human secreted pro
C 23	13.4	89.3	712	22	AAK05530	Human secreted pro
C 24	13.4	89.3	717	21	AAK46427	Arabidopsis thalia
C 25	13.4	89.3	721	21	AAK33968	Arabidopsis thalia
C 26	13.4	89.3	916	22	AAI25026	Human breast cance
C 27	13.4	89.3	1040	21	AAZ48812	Soybean inositol 1
C 28	13.4	89.3	1084	11	AAQ03671	Zucchini yellow mo
C 29	13.4	89.3	1232	21	AAK46317	Arabidopsis thalia
C 30	13.4	89.3	1234	21	AAK34736	Arabidopsis thalia
C 31	13.4	89.3	1284	22	AAK71495	Corynebacterium ql
C 32	13.4	89.3	1455	21	AAK21889	Human breast and o
C 33	13.4	89.3	1504	21	AAK32917	Arabidopsis thalia
C 34	13.4	89.3	1653	21	AAK32938	Arabidopsis thalia
C 35	13.4	89.3	1759	21	AAK47078	Arabidopsis thalia
C 36	13.4	89.3	1795	22	AAI26747	Human breast cance
C 37	13.4	89.3	1816	22	AAK16430	Human cDNA sequenc
C 38	13.4	89.3	1955	22	AAK09956	Human drug metabol
C 39	13.4	89.3	1959	21	AAA59835	Human SCIS encodi
C 40	13.4	89.3	1959	21	AAA53429	Human thiorodoxin
C 41	13.4	89.3	2056	21	AAA59836	Human thiorodoxin
C 42	13.4	89.3	2056	21	AAA53430	Human thiorodoxin
C 43	13.4	89.3	2187	21	AAZ92220	Human thiorodoxin
C 44	13.4	89.3	2203	22	AAH98122	Human EST-derived
C 45	13.4	89.3	2255	21	AAK79964	Human secreted pro

ALIGNMENTS

RESULT 1
ABL33514/c
ID ABL33514 standard; DNA; 15399 bp.

AC ABL33514;
XX
XX 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 1487.

Human; immune system disease; cytosine methylation; antiasthmatic;
antiarteriosclerotic; antianaemic; cytostatic; neurotropic;
neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;
Kw antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
Kw antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
Kw acute myeloid leukaemia; Alzheimer's disease; epilepsy;
Kw neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
gene; ds.

OS Homo sapiens.
XX
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP07537.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.

PI Olek A. Piepenbrock C, Berlin K;
XX WPI: 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
XX
XX Claim 1; SEQ ID NO 1487; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/alcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
XX
XX Sequence 15399 BP; 4553 A; 156 C; 3143 G; 7547 T; 0 other;
SQ
Query Match 100.0%; Score 15; DB 24; Length 15399;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ACAATTTCACGCA 15
Db 13487 ACAATTTCACGCA 13473
|||||||
RESULT 2
AAL24605/c
ID AAL24605 standard; cDNA; 752 BP.
XX
XX AAL24605;
XX
XX 07-DEC-2001 (first entry)
XX
XX Human breast cancer expressed polynucleotide 17062.
DE
XX
XX Human; breast cancer; cell marker; cytostatic; ss.
XX
XX Homo sapiens.
XX
XX WO200151628-A2.
PN
XX
XX 19-JUL-2001.
XX
XX 10-JAN-2001; 2001WO-US00798.
XX
XX 14-JAN-2000; 2000US-0176077.
XX 14-MAR-2000; 2000US-0189167.
XX 24-MAR-2000; 2000US-0192099.
XX 29-MAR-2000; 2000US-0193480.
XX 15-MAY-2000; 2000US-0205230.
XX 09-JUN-2000; 2000US-0211315.
XX 25-JUL-2000; 2000US-0226534.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lillie J, Xu Y, Wang Y, Steinmann K;
PI
XX
XX WPI: 2001-451856/48.
XX
XX New peptide useful as a marker for the diagnosis of breast cancer -
PT
XX
XX Claim 1; Page 3138; 3695pp; English.
XX
XX The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for

CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
XX activity.
XX
XX Sequence 752 BP; 199 A; 176 C; 197 G; 180 T; 0 other;
SQ
Query Match 93.3%; Score 14; DB 22; Length 752;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ACAATTTCACGCA 14
Db 654 ACAATTTCACGCA 641
|||||||
RESULT 3
AAL07036/c
ID AAL07036 standard; DNA; 7588 BP.
XX
XX AAL07036;
XX
XX 21-NOV-2001 (first entry)
XX
XX Human reproductive system related antigen DNA SEQ ID NO: 9724.
DE
XX
XX Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX
XX Homo sapiens.
XX
XX WO200155320-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 11-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0299287.
 PR 01-SEP-2000; 2000US-0299343.
 PR 01-SEP-2000; 2000US-0299344.
 PR 01-SEP-2000; 2000US-0299345.
 PR 05-SEP-2000; 2000US-0299509.
 PR 05-SEP-2000; 2000US-0299513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233084.
 PR 14-SEP-2000; 2000US-0233085.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-465570/50.
 XX
 PT Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition
 XX
 PS Disclosure; SEQ ID NO 9724; 1297pp + Sequence Listing; English.
 XX
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a genomic sequence encoding a
 CC protein of the invention.
 XX
 SQ Sequence 7588 BP; 1995 A; 1645 C; 1655 G; 2293 T; 0 other;

Query Match 93.3%; Score 14; DB 22; Length 7588;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CAATTTTCCACCA 15
 |||||
 DB 1084 CAATTTTCCACCA 1071

RESULT 4
 AA162714/C
 ID AA162714 standard; DNA: 7588 BP.
 XX
 AC AA162714;
 XX

DT 19-OCT-2001 (first entry)

DE Human breast or ovarian antigen genomic DNA SEQ ID NO: 364.

KW Human; breast antigen; ovarian antigen; cancer; metastasis; gene therapy;
 ds.
 XX

OS Homo sapiens.

PN WO200155324-A2.

XX

XX New isolated nucleic acids and polypeptides, useful for diagnosing,
PT treating and/or preventing human diseases and disorders -
XX
PS Disclosure; SEQ ID NO: 364; 520pp + Sequence Listing: English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of ovarian and breast antigens. These are shown in
CC A162467-AA162572 and AAM42240-AAM42345. The sequences can be used in the
CC diagnosis, prevention and treatment of breast and ovarian cancers, and
CC their metastases. The present sequence is a genomic sequence of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7588 BP; 1995 A; 1645 C; 1655 G; 2293 T; 0 other;

Query Match 93.3%; Score 14; DB 22; Length 7588;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CAATTTTCCACGA 15
|||||

Db 1084 CANTTTCCACGA 1071

RESULT 5
AAC62130
XX AAC62130 standard: DNA; 34185 BP.
AC AAC62130;
XX
DI 06-MAR-2001 (first entry)
XX
DE Nucleotide sequence of the entire bovine adenovirus type 1 genome.
XX
KW Early gene region; E3; E4; vaccine; bovine adenovirus type 1; BAV-1;
KW gene therapy; ss.
XX
OS Mastadenovirus bosl.
XX
FH Key Location/Qualifiers
FT 278..736
FT CDS /*tag= a
FT /*note= "ORF1"
FT 697..1167
FT CDS /*tag= b
FT /*note= "ORF2"
FT 1400..1867
FT CDS /*tag= c
FT /*note= "ORF1"
FT 2189..2656
FT CDS /*tag= d
FT /*note= "ORF2"
FT 2566..3777
FT CDS /*tag= e
FT /*note= "ORF3"
FT 3838..4185
FT CDS /*tag= f
FT /*note= "ORF4"
FT complement (5315..4197)
FT /*tag= g
FT /*note= "ORF5"
FT 5634..5975
FT CDS /*tag= h
FT /*note= "ORF3"
FT complement (8530..5285)
FT /*tag= i
FT /*note= "ORF6"
FT 6255..6680
FT CDS /*tag= j
FT /*note= "ORF7"
FT complement (33384..32956)

complement (10185..8527)
/*tag= k
/*note= "ORF8"
10376..11437
/*tag= l
/*note= "ORF9"
complement (10669..10301)
/*tag= m
/*note= "ORF4"
11465..13174
/*tag= n
/*note= "ORF10"
complement (13212..12607)
/*tag= o
/*note= "ORF5"
13235..14662
/*tag= p
/*note= "ORF11"
14725..15207
/*tag= q
/*note= "ORF12"
15267..16388
/*tag= r
/*note= "ORF13"
16703..17113
/*tag= s
/*note= "ORF14"
17509..20238
/*tag= t
/*note= "ORF15"
20241..20864
/*tag= u
/*note= "ORF16"
complement (22246..20906)
/*tag= v
/*note= "ORF17"
22258..24498
/*tag= w
/*note= "ORF18"
24212..24796
/*tag= x
/*note= "ORF19"
25009..25680
/*tag= y
/*note= "ORF20"
25362..25365
/*tag= b
25673..26041
/*tag= z
/*note= "ORF21"
25923..27287
/*tag= aa
/*note= "ORF22"
27291..27296
/*tag= c
27483..29294
/*tag= ab
/*note= "ORF23"
29059..29065
/*tag= d
complement (29730..29311)
/*tag= ac
/*note= "ORF24"
complement (30739..30404)
/*tag= ad
/*note= "ORF25"
complement (31464..30730)
/*tag= ae
/*note= "ORF26"
complement (32232..31471)
/*tag= af
complement (33384..32956)


```

FT CDS 27483...29294
FT /*tag= ab
FT /*note= "ORF23"
FT polyA_signal 29059...29065
FT /*tag= d
FT CDS complement (29730...29311)
FT /*tag= ac
FT /*note= "ORF24"
FT complement (30739...30404)
FT /*tag= ad
FT /*note= "ORF25"
FT CDS complement (31464...30730)
FT /*tag= ae
FT /*note= "ORF26"
FT CDS complement (32232...31471)
FT /*tag= af
FT /*note= "ORF27"
FT CDS complement (33384...32956)
FT /*tag= ag
FT /*note= "ORF28"
FT TATA_signal 34171...34174
FT /*tag= e
FT WO2000061773-A1.
XX
XX 19-OCT-2000.
XX
XX 07-APR-2000; 2000WO-US09459.
XX
XX 09-APR-1999; 99US-0289930.
XX
XX (SCHE ) SCHERING-PLOUGH LTD.
XX
XX Chiang CH, Cochran MD;
XX WPI; 2000-656327/63.
XX
XX Recombinant and mutant viruses derived from bovine adenovirus useful
XX for generating vaccines that induce an immune response in animals -
XX
XX Example 1; Page 47-57; 63pp; English.
XX
XX The present sequence represents the entire genome of a bovine
XX adenovirus type 1 (BAV-1). BAV-1 exhibits a complex sequence
XX organisation at its left and right ends. The genome exhibits an inverted
XX terminal repeat (ITR) of 578 bp. A sequence of 419 bp is repeated twice
XX at the left end of the genome. A single inverted copy of this repeat
XX occurs at the right end of the genome. Foreign DNA sequence inserted
XX into the early gene region 3 (E3) or E4 of BAV-1. The recombinant
XX adenovirus is useful for inducing an immunological response in animals.
XX The recombinant viruses are suitable as vaccines and for gene therapy.
XX
XX Sequence 34185 BP; 8803 A; 8457 C; 8236 G; 8689 T; 0 other;
XX
XX Query Match 93.3%; Score 14; DB 21; Length 34185;
XX Best Local Similarity 100.0%; Pred. NO. 3.3e+02;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 ACAATTTTCCACAG 14
XX ||||||||||||
XX Db 914 ACAATTTTCCACAG 901
XX
XX RESULT 7
XX AAL19541
XX ID AAL19541 standard; cDNA; 282 BP.
XX
XX AC AAL19541;
XX
XX 07-DEC-2001 (first entry)
XX
XX Human breast cancer expressed polynucleotide 11998.
XX

```

```

KW Human; breast cancer; cell marker; cytostatic; ss.
XX Homo sapiens.
XX WO200151628-A2.
XX
XX 19-JUL-2001.
XX
XX 10-JAN-2001; 2001WO-US00798.
XX
XX 14-JAN-2000; 2000US-0176077.
XX 14-MAR-2000; 2000US-0189167.
XX 24-MAR-2000; 2000US-0192099.
XX 29-MAR-2000; 2000US-0193480.
XX 15-MAY-2000; 2000US-0205230.
XX 09-JUN-2000; 2000US-0211315.
XX 25-JUL-2000; 2000US-0220534.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX
XX New peptide useful as a marker for the diagnosis of breast cancer
XX
XX Claim 1; Page 2123; 3695pp; English.
XX
XX The invention relates to human breast cancer expressed polynucleotides
XX (AAL07544-AAL26789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterising treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cytostatic
XX activity.
XX
XX Sequence 282 BP; 83 A; 68 C; 57 G; 74 T; 0 other;
XX
XX Query Match 89.3%; Score 13.4; DB 22; Length 282;
XX Best Local Similarity 93.3%; Pred. NO. 5.1e+02;
XX Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 ACAATTTTCCACCA 15
XX ||||||||||||
XX Db 244 ACAATTTTCCACCA 258
XX
XX RESULT 8
XX AAF66087
XX ID AAF66087 standard; cDNA; 380 BP.
XX
XX AC AAF66087;
XX
XX DT 09-APR-2001 (first entry)
XX
XX DE Novel human polynucleotide, SEQ ID NO: 1843.
XX
XX Human; cytostatic; gene therapy; colon cancer; prostate cancer;
XX breast cancer; lung cancer; cancer detection; ss.
XX
XX Homo sapiens.
XX
XX WO200102568-A2.
XX
XX 11-JAN-2001.
XX
XX 30-JUN-2000; 2000WO-US18374.
XX
XX 02-JUL-1999; 99US-0142310.
XX 02-JUL-1999; 99US-0142311.
XX

```


DR WPI; 2001-483447/52.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 PT
 XX
 PS Claim 1; SEQ ID NO 4885; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX

30 Sequence 450 BP; 120 A; 114 C; 87 G; 129 T; 0 other;
 Query Match 89.3%; Score 13.4; DB 22; Length 450;
 Best Local Similarity 93.3%; Pred. No. 5.2e+02;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAATTTTCCACGA 15
 ||||| |||||
 Db 342 ACAATTTTCCACGA 328

RESULT 11

ABA26195/C
 ID ABA26195 standard; DNA; 450 BP.
 XX
 AC ABA26195;
 XX
 XX
 DT 23-JAN-2002 (first entry)
 XX
 DE Probe #4661 for gene expression analysis in human heart cell sample.
 XX
 XX Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200157274-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00666.

PF 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human hearts -

XX Claim 1; SEQ ID No 4661; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived

CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

SQ Sequence 450 BP; 120 A; 114 C; 87 G; 129 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22;

Best Local Similarity 93.3%; Pred. No. 5.2e+02;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAATTTTCCACGA 15
 ||||| |||||

Db 342 ACAATTTTCCACGA 328

RESULT 12

AAK04728/C

ID AAK04728 standard; DNA; 450 BP.

XX AAK04728;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe SEQ ID NO: 4719.

XX Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer; ss.

XX Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human brains -

XX Example 4; SEQ ID NO: 4719; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention.

XX Sequence 450 BP; 120 A; 114 C; 87 G; 129 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22;

Best Local Similarity 93.3%; Pred. No. 5.2e+02;


```

PF 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta.
XX
XX Claim 25; SEQ ID No 4893; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 450 BP: 120 A; 114 C; 87 G; 129 T; 0 other;
SQ
Query Match 89.38; Score 13.4; DB 22; Length 450;
Best Local Similarity 93.38; Pred. No. 5.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ACAATTTTCCACGA 15
Db 342 ACAATTTTCCACGA 328

```

Search completed: October 17, 2002, 10:09:26
Job time : 168 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 08:51:00 : Search time 32 Seconds
(without alignments)
115.141 Million cell updates/sec

Title: US-09-820-203A-2
Perfect score: 15
Sequence: 1 acaatttcccacga 15

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2.6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2.6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2.6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2.6/ptodata/1/ina/PTGUS_COMB.seq.*
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	13	86.7	682	4	US-09-129-030-51
C 2	12.4	82.7	21	1	US-08-064-271-14
C 3	12.4	82.7	21	3	US-08-930-589A-12
C 4	12.4	82.7	27	1	US-08-064-271-3
C 5	12.4	82.7	27	3	US-08-712-610-3
C 6	12.4	82.7	27	3	US-08-930-589A-3
C 7	12.4	82.7	27	4	US-09-309-423-3
C 8	12.4	82.7	200	2	US-08-875-972-15
C 9	12.4	82.7	219	1	US-08-702-080-3
C 10	12.4	82.7	219	2	US-08-858-830-3
C 11	12.4	82.7	219	2	US-08-858-834-3
C 12	12.4	82.7	590	4	US-09-129-030-1
C 13	12.4	82.7	795	3	US-08-904-284-2
C 14	12.4	82.7	1407	3	US-08-688-988-7
C 15	12.4	82.7	1539	3	US-09-058-725R-6
C 16	12.4	82.7	1539	3	US-09-232-857-6
C 17	12.4	82.7	2348	2	US-08-876-546A-1
C 18	12.4	82.7	2348	4	US-09-412-252-1
C 19	12.4	82.7	2359	1	US-08-188-582-4
C 20	12.4	82.7	2359	1	US-08-646-715-4
C 21	12.4	82.7	2679	3	US-08-904-284-4
C 22	12.4	82.7	2774	3	US-08-732-433-2
C 23	12.4	82.7	3387	2	US-08-627-254C-29
C 24	12.4	82.7	5105	1	US-08-148-122A-1
C 25	12.4	82.7	5362	3	US-08-463-210-5
C 26	12.4	82.7	7399	2	US-08-418-848A-9
C 27	12.4	82.7	9709	2	US-08-188-583-5

C 28	12.4	82.7	9709	3	US-08-388-353-1
C 29	12.4	82.7	9709	3	US-08-488-551B-1
C 30	12.4	82.7	9737	2	US-08-944-449-7
C 31	12.4	82.7	9739	1	US-08-022-835-1
C 32	12.4	82.7	9739	1	US-08-388-809-1
C 33	12.4	82.7	979	2	US-08-647-714-1
C 34	12.4	82.7	979	2	US-08-022-835-3
C 35	12.4	82.7	979	2	US-08-388-809-3
C 36	12.4	82.7	9745	2	US-08-647-714-3
C 37	12.4	82.7	12494	4	US-08-935-312-13
C 38	12.4	82.7	12494	4	US-08-848-760B-33
C 39	12.4	82.7	15581	3	US-08-646-538-35
C 40	12.4	82.7	15581	4	US-09-503-222-35
C 41	12.4	82.7	35524	3	US-08-923-137-1
C 42	12.4	82.7	36519	3	US-08-923-137-2
C 43	12	80.0	413	2	US-08-980-060-11
C 44	12	80.0	413	4	US-09-307-185-11
C 45	12	80.0	642	3	US-08-945-026-33

ALIGNMENTS

RESULT 1
US-09-129-030-51/c
; Sequence 51, Application US/09129030A
; Patent No. 6242221

; GENERAL INFORMATION:
; APPLICANT: COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION
; TITLE OF INVENTION: GENOMIC PPO CLONES
; FILE REFERENCE: 57072-PCT-US
; CURRENT APPLICATION NUMBER: US/09/129,030A
; CURRENT FILING DATE: 1998-08-04
; EARLIER APPLICATION NUMBER: AU PN7856
; EARLIER FILING DATE: 1996-02-05
; EARLIER APPLICATION NUMBER: AU P02361
; EARLIER FILING DATE: 1996-09-16
; EARLIER APPLICATION NUMBER: PCT/AU97/00041
; EARLIER FILING DATE: 1997-01-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 682
; TYPE: DNA
; ORGANISM: POTATO
; FEATURE:
; NAME/KEY: QDS
; LOCATION: (3)...(680)
US-09-129-030-51

Query Match 86.7%; Score 13; DB 4; Length 682;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AATTTCACCA 15
Db 127 AATTTCACCA 115

RESULT 2
US-08-064-271-14/c
; Sequence 14, Application US/08064271
; Patent No. 5543297

; GENERAL INFORMATION:
; APPLICANT: Kennedy, Brian P.
; APPLICANT: Cromlish, Wanda A.
; APPLICANT: Mancini, Joseph A.
; APPLICANT: O'Neil, Gary
; APPLICANT: Vickers, Philip J.
; APPLICANT: Wong, Elizabeth
; TITLE OF INVENTION: HUMAN CYCLOOXYGENASE-2 CDNA AND
; TITLE OF INVENTION: ASSAY FOR EVALUATING CYCLOOXYGENASE ACTIVITY
; NUMBER OF SEQUENCES: 14

```

: REGISTRATION NUMBER: 38,413
: REFERENCE/DOCKET NUMBER: 19029PC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 732-594-6734
: TELEFAX: 732-594-4720
: TELEX:
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 21 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-930-589A-12

Query Match      82.7%; Score 12.4; DB 3; Length 21;
Best Local Similarity 92.9%; Pred. No. 1.5e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CAATTTCACGCA 15
      IIIIIIIIIII
Db      18 CAATTTCACGCA 5

RESULT 4
US-08-064-271-3/c
: Sequence 3, Application US/08064271
: Patent No. 5543297
: GENERAL INFORMATION:
: APPLICANT: Kennedy, Brian P.
: APPLICANT: Cromlish, Wanda A.
: APPLICANT: Mancini, Joseph A.
: APPLICANT: O'Neil, Gary
: APPLICANT: Vickers, Philip J.
: APPLICANT: Wong, Elizabeth
: TITLE OF INVENTION: HUMAN CYCLOOXYGENASE-2 cDNA AND
: TITLE OF INVENTION: ASSAY FOR EVALUATING CYCLOOXYGENASE ACTIVITY
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Merck & Co., Inc.
: STREET: 126 Lincoln Avenue
: CITY: Rahway
: STATE: NJ
: COUNTRY: USA
: ZIP: 07065
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.5 in, 1.4kb
: COMPUTER: Apple Macintosh
: OPERATING SYSTEM: System 7
: SOFTWARE: Microsoft Word 5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/064,271
: FILING DATE: 19930506
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Panzer, Curtis C.
: REGISTRATION NUMBER: 33,752
: REFERENCE/DOCKET NUMBER: 189061A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (908)594-3199
: TELEFAX: (908)594-4720
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 27 bases
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-064-271-3

Query Match      82.7%; Score 12.4; DB 1; Length 27;
Best Local Similarity 92.9%; Pred. No. 1.5e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 2 CAATTTTCCACGA 15
| | | | | | | | | |
Db 25 CAATTTTCCACAA 12

RESULT 5

US-08-712-610-3/c
; Sequence 3, Application US/08712610
; Patent No. 6048897
; GENERAL INFORMATION:
; APPLICANT: Charles N. Serhan
; TITLE OF INVENTION: Lipoxin Compounds
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/712,610
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/260,030
; FILING DATE: 15-JUNE-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,300
; FILING DATE: 15-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: BWI-112CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; ANTI-SENSE: NO
US-08-712-610-3

Query Match 82.7%; Score 12.4; DB 3; length 27;
Best Local Similarity 92.9%; Pred. No. 1.5e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CAATTTTCCACGA 15
| | | | | | | | | |
Db 25 CAATTTTCCACAA 12

RESULT 6

US-08-930-589A-3/c
; Sequence 3, Application US/08930589A
; Patent No. 6107087
; GENERAL INFORMATION:
; APPLICANT: MERCK FROSST CANADA & CO.
; APPLICANT: O'NEILL, GARY P.
; TITLE OF INVENTION: HIGH LEVEL EXPRESSION OF HUMAN
; CYCLOOXYGENASE-2
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, 126 E. Lincoln Ave.
; CITY: Rahway
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/930,589A
; FILING DATE: 28-JUN-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coppola, Joseph A.
; REGISTRATION NUMBER: 38,413
; REFERENCE/DOCKET NUMBER: 19029PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-6734
; TELEFAX: 732-594-4720
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-930-589A-3

Query Match 82.7%; Score 12.4; DB 3; length 27;
Best Local Similarity 92.9%; Pred. No. 1.5e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CAATTTTCCACGA 15
| | | | | | | | | |
Db 25 CAATTTTCCACAA 12

RESULT 7

US-09-309-423-3/c
; Sequence 3, Application US/09309423
; Patent No. 6316648
; GENERAL INFORMATION:
; APPLICANT: Charles N. Serhan
; TITLE OF INVENTION: Lipoxin Compounds
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/309,423
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/260,030
; FILING DATE: 15-JUNE-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,300
; FILING DATE: 15-JUNE-1993


```

; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: BWI-112CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; ANTI-SENSE: NO
; US-09-309-423-3

Query Match      82.7%; Score 12.4; DB 4; Length 27;
Best Local Similarity 92.9%; Pred. No. 1.5e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CAATTTTCCACGA 15
        |||||
        25 CAATTTTCCACAA 12

RESULT 8
US-08-875-972-15
; Sequence 15, Application US/08875972
; Patent No. 5985564
; GENERAL INFORMATION:
; APPLICANT: Huntington Potter and Jirihue Li
; TITLE OF INVENTION: ASSAY FOR IDENTIFYING GENES CAUSING
; TITLE OF INVENTION: CHROMOSOME NON-DISJUNCTION
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/875,972
; FILING DATE: 08-AUG-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/002,448
; FILING DATE: 16-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan Esq., Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: HU95-03PA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 861-6240
; TELEFAX: (781) 861-9540
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-875-972-15

Query Match      82.7%; Score 12.4; DB 2; Length 200;
Best Local Similarity 86.7%; Pred. No. 1.9e+02;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: BWI-112CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; ANTI-SENSE: NO
; US-09-309-423-3

Query Match      82.7%; Score 12.4; DB 1; Length 219;
Best Local Similarity 92.9%; Pred. No. 1.9e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ACAATTTTCCACGA 15
        |||||
        55 ACAATTTTCCACGA 69

RESULT 9
US-08-702-080-3
; Sequence 3, Application US/08702080
; Patent No. 5654173
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavalie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/702,080
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-702-080-3

Query Match      82.7%; Score 12.4; DB 1; Length 219;
Best Local Similarity 92.9%; Pred. No. 1.9e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CAATTTTCCACGA 15
        |||||
        43 CAATTTTCCACGA 56

RESULT 10
US-08-858-830-3
; Sequence 3, Application US/08858830
; Patent No. 5965693
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavalie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
```

```
;
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858-830
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/702,080
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-858-830-3

Query Match      82.7%  Score 12.4; DB 2; Length 219;
Best Local Similarity 92.9%; Pred. No. 1.9e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2 CAATTTCCACGA 15
    |||||
Db   43 CAATTTCCAAGA 56

RESULT 11
US-08-858-834-3
; Sequence 3, Application US/08858834
; Patent No. 5981222
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: LaVallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,834
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/702,080
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-858-834-3

Query Match      82.7%  Score 12.4; DB 2; Length 219;
Best Local Similarity 92.9%; Pred. No. 1.9e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2 CAATTTCCACGA 15
    |||||
Db   43 CAATTTCCAAGA 56

RESULT 12
US-09-129-030-1/c
; Sequence 1, Application US/09129030A
; Patent No. 6242221
; GENERAL INFORMATION:
; APPLICANT: COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION
; TITLE OF INVENTION: GENOMIC PRO CLONES
; FILE REFERENCE: 57072-PCT-US
; CURRENT APPLICATION NUMBER: US/09/129,030A
; CURRENT FILING DATE: 1998-08-04
; EARLIER APPLICATION NUMBER: AU PN7856
; EARLIER FILING DATE: 1996-02-05
; EARLIER APPLICATION NUMBER: AU P02361
; EARLIER FILING DATE: 1996-09-16
; EARLIER APPLICATION NUMBER: PCT/AU97/00041
; EARLIER FILING DATE: 1997-01-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 590
; TYPE: DNA
; ORGANISM: STRAWBERRY
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(588)
; US 09-129-030-1

Query Match      82.7%  Score 12.4; DB 4; Length 590;
Best Local Similarity 92.9%; Pred. No. 2.1e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2 CAATTTCCACGA 15
    |||||
Db   57 CAATTTCCACGA 44

RESULT 13
US-08-904-284-2/c
; Sequence 2, Application US/08904284
; Patent No. 6133435
; GENERAL INFORMATION:
; APPLICANT: Fernandez, Donna E.
```

```

/ APPLICANT: Heck, Gregory R.
/ TITLE OF INVENTION: EXPRESSION OF AGL15 SEQUENCE IN
/ NUMBER OF SEQUENCES: 7
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Quarles & Brady
/ STREET: 1 South Pinckney Street
/ CITY: Madison
/ STATE: WI
/ COUNTRY: US
/ ZIP: 53701-2113
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/904,284
/ FILING DATE:
/ CLASSIFICATION: 800
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Seay, Nicholas J.
/ REGISTRATION NUMBER: 27,386
/ REFERENCE/DOCKET NUMBER: 960296,94193
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (608) 251-5000
/ TELEFAX: 608-251-9166
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 795 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ US-08-904-284-2

```

```

Query Match      82.7%; Score 12.4; DB 3; Length 795;
Best Local Similarity 92.9%; Pred. No. 2.1e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 2 CAATTTTCCACGA 15
    ||||| |||||
Db 19 CAATTTTCCACGA 6

```

```

RESULT 14
US-08-688-988-7
/ Sequence 7, Application US/08688988B
/ Patent No. 6096545
/ GENERAL INFORMATION:
/ APPLICANT: Lefebvre, Daniel D.
/ APPLICANT: Malboobi, Mohammad A.
/ TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS
/ FILE REFERENCE: PPL96-03
/ CURRENT APPLICATION NUMBER: US/08/688,988B
/ CURRENT FILING DATE: 1996-07-31
/ NUMBER OF SEQ ID NOS: 48
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 7
/ LENGTH: 1407
/ TYPE: DNA
/ ORGANISM: Brassica Nigra
/ US-08-688-988-7

```

```

Query Match      82.7%; Score 12.4; DB 3; Length 1407;
Best Local Similarity 92.9%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 2 CAATTTTCCACGA 15
    ||||| |||||
Db 144 CAATTTTCCACGA 157

```

```

RESULT 15
US-09-058-725R-6
/ Sequence 6, Application US/09058725B
/ Patent No. 6133420
/ GENERAL INFORMATION:
/ APPLICANT: Ames, Robert
/ APPLICANT: Sarau, Henry
/ APPLICANT: Foley, James
/ APPLICANT: Chamber, Jon
/ TITLE OF INVENTION: A Method of Finding Angonist
/ TITLE OF INVENTION: and Antagonist To Human and Rat GPR 14
/ NUMBER OF SEQUENCES: 6
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SmithKline Beecham Corporation
/ STREET: 709 Swedeland Road
/ CITY: King of Prussia
/ STATE: PA
/ COUNTRY: USA
/ ZIP: 19406
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/058,725H
/ FILING DATE: April 10, 1998
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/789,354
/ FILING DATE: 27-JAN-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Han, William T.
/ REGISTRATION NUMBER: 34,344
/ REFERENCE/DOCKET NUMBER: GP50005-1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 610-270-5219
/ TELEFAX: 610-270-5090
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1539 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ US-09-058-725R-6

```

```

Query Match      82.7%; Score 12.4; DB 3; Length 1539;
Best Local Similarity 92.9%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 ACAATTTCCACG 14
    ||||| |||||
Db 212 ACAATTTCCACG 225

```

```

Search completed: October 17, 2002, 10:11:51
Job time : 35 secs

```

GenCore version 5.1.3
Copyright (C) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 09:32:18 ; Search time 1196 Seconds
(without alignments)
169.276 Million cell updates/sec

Title: US-09-820-203A-2

Perfect score: 15

Sequence: 1 acaatttccacga 15

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 274724 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpi:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	100.0	483	B1434795	B1434795 EST537556
2	15	100.0	549	BH023732	BH023732 BG01382 D
3	15	100.0	578	FR0045169	AL139458 Fugu rubr
4	15	100.0	583	BF245266	BF245266 601863739
5	15	100.0	621	A2489193	A2489193 LM0319C21
6	15	100.0	648	AI208513	AI208513 q956h01.x
7	15	100.0	770	BH202289	BH202289 Sml-55N13
8	15	100.0	849	BH184804	BH184804 026_M_09-
9	15	100.0	849	CNS07PNT	AI621755 t7 end of
10	14	93.3	224	BH314158	BH314158 ih07h07.y
11	14	93.3	283	AQ017178	AQ017178 CIT-HSP-2
12	14	93.3	300	C07471	C07471 C07471 fuji
13	14	93.3	357	9	BB332438 BB332438
14	14	93.3	363	10	BF730146 BF730146 mab70c04
15	14	93.3	426	12	AQ056038 CIT-HSP-2
16	14	93.3	439	10	BM313878 BM313878
17	14	93.3	430	12	A2072907 RPI-23-4

18 14 93.3 450 10 RF117715
19 14 93.3 454 12 AQ232160
20 14 93.3 532 9 AW494391
21 14 93.3 550 12 B65275
22 14 93.3 639 9 BB658770
23 14 93.3 699 10 BF532786
24 14 93.3 714 12 A2176016
25 14 93.3 733 12 AG011330
26 14 93.3 813 10 BG758425
27 14 93.3 837 10 BG165717
28 14 93.3 909 10 BF338620
29 14 93.3 913 10 BF967150
30 14 93.3 925 10 BF164443
31 14 93.3 943 10 B1329624
32 14 93.3 1012 10 BG323232
33 14 93.3 1041 10 BM454453
34 14 93.3 1101 12 CNS00F55
35 14 93.3 1110 10 BF310655
36 14 93.3 1128 10 BG622546
37 14 93.3 1205 12 B09441
38 13.4 89.3 144 9 AW112155
39 13.4 89.3 165 9 AA804244
40 13.4 89.3 167 10 BF318671
41 13.4 89.3 170 9 BB109174
42 13.4 89.3 171 9 AA883025
43 13.4 89.3 172 10 BF345067
44 13.4 89.3 180 10 C30296
45 13.4 89.3 195 10 B1358314

ALIGNMENTS

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone
PPCB070 5' sequence, mRNA sequence.

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

B1434795 483 bp mRNA linear EST 21-AUG-2001
EST537556 P. infestans-challenged leaf Solanum tuberosum cDNA clone

no signs of HR. katahdin plants (susceptible to P. infestans US-1) were used as controls and showed infection. NOTE: We cannot exclude the possibility that this sequence is actually derived from Phytophthora rather than potato."

BASE COUNT 135 a 94 c 91 g 163 t

Query Match 100.0%; Score 15; DB 10; Length 483;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAATTTTCCACGA 15
|||||

Db 378 ACAATTTTCCACGA 392

RESULT 2

BH023732

LOCUS

DEFINITION BH023732 549 bp DNA linear GSS 09-JUL-2001
Drosophila melanogaster P[GTI] P element insertion lines
3' ends of P element, DNA sequence.

ACCESSION BH023732

VERSION BH023732.1 GI:14627191

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 549)

AUTHORS Lewis, R., Hoskins, R., Liao, G., Morzen, N., Tsang, G., He, Y., Karpen

G., Belen, H., Rubin, G., and Spradling, A.

The Berkeley Drosophila Genome Project

Unpublished (2001)

Contact: Gerald Rubin

Berkeley Drosophila Genome Project

University of California, Berkeley

LSA Building, Berkeley, CA 94720-3200, USA

Fax: 5106439947

Email: gerry@fruitfly.berkeley.edu

Sequence recovery method was inverse PCR.

Sequence orientation is forward strand relative to 5' end of P

element

The P element insertion position is base 486 in the 549 bases. This

insertion position refers to the first base of the 8 base target

recognition sequence.

Class: transposon-tagged.

Location/Qualifiers

1..549

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="Drosophila melanogaster P[GTI] P element

insertion lines"

/note="Inverse PCR was performed on Drosophila

melanogaster strains each of which contains one or more

P[GTI] P-element transposon insertion. The resultant

fragment for each strain was directly sequenced to

determine the genomic sequence at the site of insertion.

Details of the protocols used can be found at

http://www.fruitfly.org/about/methods/inverse.pcr.html."

BASE COUNT 180 a 99 c 91 g 179 t

ORIGIN

Query Match 100.0%; Score 15; DB 12; Length 549;

Best Local Similarity 100.0%; Pred. No. 1e+03;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAATTTTCCACGA 15

|||||

Db 299 ACAATTTTCCACGA 313

RESULT 3

FR0045169/c

LOCUS

DEFINITION FR0045169 578 bp DNA linear GSS 12-JUN-2000
Fugu rubripes GSS sequence, clone 061D18aF8, genomic survey

ACCESSION AL139458

VERSION AL139458.1 GI:8568458

KEYWORDS GSS; genome survey sequence.

SOURCE Takifugu rubripes.

ORGANISM Takifugu rubripes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontidae; Takifugu.

REFERENCE 1 (bases 1 to 578)

Power, D.

Direct Submission

Submitted (15-FEB-2000) MRC Human Genome Mapping Project Resource

Centre Hinxton, Cambridge, CB10 1SB, UK Email:

b.loheip@hmp.mrc.ac.uk

Vector: pBluescript II KS

V-type: phagemid

PRIMER: KS

DESCR:

One pass dye-terminator sequencing of cosmid cloned genomic

sequence.

Location/Qualifiers

1..578

/organism="Takifugu rubripes"

/db_xref="taxon:31033"

/clone_lib="cosmid 061D18"

/clone="061D18aF8"

BASE COUNT 130 a 133 c 150 g 137 t 28 others

ORIGIN

Query Match 100.0%; Score 15; DB 12; Length 578;

Best Local Similarity 100.0%; Pred. No. 1e+03;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAATTTTCCACGA 15

|||||

Db 389 ACAATTTTCCACGA 375

RESULT 4

BF245266

LOCUS

DEFINITION BF245266 583 bp mRNA linear EST 14-NOV-2000
601863739F1 NI⁺ MGC_57 Homo sapiens CDNA clone IMAGE:4081352 5',

ACCESSION BF245266

VERSION BF245266.1 GI:11159198

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 583)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: CLONETECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LICM942 row: n column: 09

High quality sequence stop: 141.

Location/Qualifiers

```

FEATURES
    source
        nridn quality sequence stop: 273.
        Location/Qualifiers
            1. .648
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:1839217"
                /clone_lib="Scares_testis_NHT"
                /sex="male"
                /lab_host="DH10B"
                /note="Vector: pT73b-Pac (pharmacia) with a modified
                polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA
                was prepared from mRNA obtained from Clontech laboratories
                , inc., and primed with a Not I - oligo(dT) primer [5,
                TGCTACCAATCTGAAGTGGAGCGCGGCCCAATTTTTTTTTTTTTT 3'].
                Double-stranded cDNA was ligated to Eco RI adaptors.

```

```

Query Match      100.0%  Score 15;  DB 12;  Length 770;
Best local Similarity 100.0%;  Pred. No. 1.1e+03;
Matches 15;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Y      1  ACAATTTTCCACGA 15
      | | | | | | | | | |
      422 ACAATTTTCCACGA 408

```

RESULT 9	CNS07PNT	849 bp	DNA	linear	GSS 10-0CT-2001
	OCUS	T7 end of clone 026AG05	of library SmBAC1	from strain Puerto-Rican	
	DEFINITION	AL621755	of Schistosoma mansoni	genomic survey sequence.	
	ACCESSION	AL621755			
	VERSION	AL621755.1	GI:16035897		
	KEYWORDS	GSS.			
	SOURCE	Schistosoma mansoni.			
	ORGANISM	Schistosoma mansoni			

